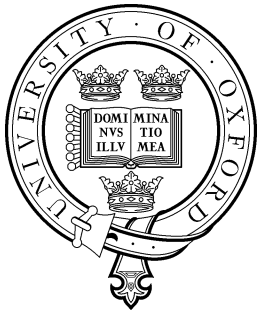
Mathematical Modelling of the Host-Virus Interaction in Chronic HTLV-I Infection and its Impact on our Understanding of Viral Persistence and Pathogenesis



Aaron Guanliang Lim
Wolfson College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary Term 2015

To my family, who have always given me the strength and courage to keep moving forward.

Acknowledgements

First and foremost, I am grateful for my supervisors, Professor Philip Maini and Professor Sunetra Gupta, whose guidance and encouragement in my DPhil project has helped me strive to do my best and reach for success. I have learnt so much from both of them and it is because of their support and dedication that I have been able to reach where I am now. I am especially thankful to Philip for his unparallelled dedication, excellent advice, and lively sense of humour — Thank you very much for being such an extraordinary mentor throughout my DPhil programme!

I would like to thank all of the members of the WCMB and the EEID for creating an invigorating atmosphere conducive to learning and research. Here's to great times and frequent tea trains with officemates, both past and present (in no particular order). [WCMB]: Jessica McGillen, Alhaji Cherif, Louise Dyson, Suruchi Bakshi, Gabriel (Gabs) Rosser, Abdullah Al-Shammari, Shuohao (Jackie) Liao, Jochen Kursawe, Linus Schumacher, Paul ('KOBD') Taylor, as well as honorary officemate, Deborah (Debbie) Markham. [EEID]: José Lourenço, Paul Wikramaratna, Ben Ashby, Eleanor (Ellie) Watkins, Robert Noble, Andrew (Andy) Walker, Farania Rangkuti, Carinna Hockham, and Susan (Suze) Hawkins.

I am deeply indebted to my previous supervisor and friend, Professor Michael Li, not only for his dedicated mentorship during my undergraduate research projects and my M.Sc. as well as introducing me to my current topic of research but, significantly, for his continued support and advice that has extended long after my time at the University of Alberta. I would also like to acknowledge Dr. Zhisheng Shuai, who helped strengthen the global stability result in Chapter 3.

I am grateful towards my DPhil examiners, Professor Eamonn Gaffney and Dr Becca Asquith, whose special care and attention in reading my thesis led to a fruitful and insightful discussion about my research project.

I would like to express my heartfelt gratitude to my family for their never-ending support and encouragement throughout the years: my father, Alan Lim, my mother, Cynthia Leow, my two brothers, Adrian and Adam, and my fiancée, Breanne Landry. A special thanks goes out to my family away from home, Yiru Lim, my bona fide friend, Hitomi Inoue, my DPhil comrade, Chris Menelaou, and my childhood buddy, Madison Adam. I could not have made it this far without being surrounded by a great support network and, although not all listed here, I am sincerely grateful to each and every one who believed in me and whose friendships I cherish dearly.

Lastly, I would like to acknowledge that this work was generously supported in part by scholarships and grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Alberta Heritage Scholarship Fund (AHSF), the Canadian Centennial Scholarship Fund (CCSF), and the Vice-Chancellors' Fund (VCF).

Mathematical Modelling of the Host-Virus Interaction in Chronic HTLV-I Infection and its Impact on our Understanding of Viral Persistence and Pathogenesis

Aaron Guanliang Lim
Wolfson College
University of Oxford
A thesis submitted for the degree of Doctor of Philosophy
Hilary Term 2015

Abstract

Human T-cell lymphotropic virus type I (HTLV-I) is a persistent human retrovirus characterised by life-long infection and risk of developing one of two major, clinically independent diseases: adult T-cell leukaemia/lymphoma (ATL), an aggressive blood cancer, and HAM/TSP, a progressive neurological and inflammatory disease. Infected individuals typically mount a large, chronically activated $CD8^+$ cytotoxic T-lymphocyte (CTL) response against HTLV-I-infected cells, but ultimately fail to effectively eliminate the virus. Moreover, the identification of determinants to disease manifestation has thus far been elusive.

A central issue in current HTLV-I research is how the virus is able to persist despite strong immune pressure. To explore this issue, we adopt a mathematical modelling approach. Mathematical modelling can help us break apart the complex mechanisms of HTLV-I persistence and identify the underlying principles that govern successful viral propagation in the presence of host immunity. Understanding these interactions is a crucial step on the road to developing effective ways to disrupt the virus life-cycle and may help identify promising new treatment strategies to reduce the severity of HTLV-I infection and minimise the detriment due to associated disease.

The objective of this thesis is to develop a consistent theoretical framework that can help shed light on specific, biologically relevant questions that are of interest to experimentalists and theoretical immunologists trying to understand the complicated host-pathogen dynamics of chronic infection by HTLV-I. In this thesis, we construct a series of mathematical models, each incorporating an increasing level of immunological detail, designed to explore the impact of three biologically significant features that have not been fully considered in previous mathematical models of HTLV-I: (I) the trade-off between proviral latency and activation, (II) the simultaneous expression of viral proteins, and (III) the role of antigenic variability.

The results from investigation of these features in the various models help contribute to our understanding of the dynamics of persistent HTLV-I infection alongside virus-specific host immunity. Although our approach is primarily theoretical in nature, our research is principally driven by a desire to elucidate biologically and clinically relevant phenomena observed in HTLV-I and, for each of the models, we link our results to what is known from experimental observations and discuss their significance in a biologically meaningful context.

Contents

Glossary/Table of Terminology **xi**

1 Introduction **1**

1.1 HTLV-I Infection 1

1.2 Mathematical Methodology 4

1.3 Objectives and Overview 4

2 HTLV-I **8**

2.1 Epidemiology 8

2.2 Classification and Viral Dynamics 10

2.3 HTLV-I Proteins 12

2.4 HTLV-I-Specific Host Immunity 14

2.5 Pathogenesis of HTLV-I-Associated Disease 15

2.6 Simultaneous Expression of Multiple Viral Proteins: Focussing on the Roles of Tax and HBZ 17

2.7 Previous Modelling Work 17

2.7.1 Wodarz, Nowak, and Bangham (1999) 17

2.7.2 Asquith and Bangham (2000) 19

2.7.3 Gómez-Acevedo, Li, and Jacobson (2010) 20

2.7.4 Li and Lim (2011) 22

3 Constructing a Baseline Mathematical Model **25**

3.1 Motivation 26

3.2 Baseline Mathematical Model Formulation 27

3.3 Parameter Values and Numerical Simulations 31

3.3.1 Parameter Values 31

3.3.2 Numerical Simulations 34

3.4 Global Qualitative Behaviour 34

3.4.1 Biologically Realistic Region 35

3.4.2 Basic Reproduction Number for Viral Infection 37
3.4.3 Characterisation of Global Dynamics 37
3.4.4 Proof of Theorem 3.2. Part 1: Existence of Equilibria 38
3.4.5 Proof of Theorem 3.2. Part 2: Stability of P_0 39
3.4.6 Proof of Theorem 3.2. Part 3: Stability of P^* 42
3.5 Non-Dimensionalisation and Asymptotic Analysis 43
3.5.1 Non-dimensionalisation 43
3.5.2 Re-scaling Variables to Order 1 44
3.5.3 Re-scaling Parameter Values to Order 1 45
3.5.4 Asymptotic Analysis 49
3.6 Measures of Viral Burden 53
3.7 Exploring Spontaneous Viral Expression: Why is HTLV-I not silent? (Question 1.) 54
3.7.1 Tax/HBZ Expression Drives Chronic Infection and is an Important Determinant of Proviral Load 55
3.7.2 Tax/HBZ Expression Increases Activated Fraction of Infected Cells . 56
3.8 Evaluating CTL Quality: What determines the strength of the HTLV-I-specific CTL response, and why is infection life-long? (Question 2.) 57
3.8.1 CTL Frequency $\neq$ CTL Efficiency 57
3.8.2 Efficient CTLs Cannot Eliminate the Active Proportion of Proviral Cells 63
3.9 Insights to HTLV-I-Associated Disease: What characterises HTLV-I-linked pathology? (Question 3.) 63
3.9.1 A Hypothesis for Determining Clinical Status 64
3.9.2 Failure of HAM/TSP Treatment 67
3.9.3 Immune Compromise is a Possible, Though Unlikely, Cause of Disease 69
3.9.4 A Plausible Mechanism for Pathogenesis 70
3.10 Summary of Results from Baseline HTLV-I Model 73
3.10.1 Mathematical Conclusions 73
3.10.2 Biological Conclusions 74
3.11 Modelling Development and Structural Stability 77
3.12 Final Remarks on the Baseline Model 78

4 Considering Simultaneous Expression of Two Epitopes in HTLV-I Infection: The Tax/HBZ Model 79

4.1 Motivation 80

4.2 Extension of Baseline Model to a Basic Multiple Epitope Model: The Tax/HBZ Model 81

4.2.1 Tax/HBZ Model Formulation 81

4.2.2 Virus Fitness and Heterogeneity in CTL Responses 85

4.2.3 A Detailed Look at the Activation and Effector Phases of the Cell-Mediated Immune Response 86

4.2.4 Global Dynamical Behaviour of the Tax/HBZ Model 88

4.3 Differences in the Effective Expansion Rate and Functionality Between Epitope-Specific CTL Responses Determine How Immunogenicity Affects Measures of CTL and Virus Abundance 89

4.3.1 CTL Dominance and CTL-to-Virus Ratio are Dependent on Immunogenicity and Factors Related to CTL Proliferation and Lifespan, but not Efficiency 89

4.3.2 The Relationship Between Proviral Load and CTL-Related Factors, Namely the Effective CTL Ratio and Immunogenicity 92

4.3.3 The Impact of the Efficiency of Each Antigen-Specific CTL Type Depends on Immunogenicity 95

4.4 Discussion of Main Results from the Tax/HBZ Model 97

4.5 Final Remarks on the Two-Epitope Tax/HBZ Model 99

5 Principles of Multiple Epitope Systems Without Antigenic Variation 100

5.1 Motivation: What is Immunodominance and Why is it Important for us to Consider? 101

5.2 Extending the Tax/HBZ Model to Incorporate an Arbitrary Number of Distinct Epitopes 102

5.2.1 Formulating a General Multiple Epitope Model of HTLV-I Infection 104

5.2.2 Global Dynamical Behaviour of the Multiple Epitope Model 105

5.2.3 Quantifying Measures of CTL Abundance: Establishing an Immunodominance Hierarchy 106

5.2.4 Constructing a Visual Representation of an Individual's CTL Response 108

5.3 Characterising the Anti-HTLV-I Cellular Immune Response Against Multiple Epitopes: Establishing a CTL Profile and Observing its Effect 113

5.3.1 The Immunogenicity Profile Alone Can Give Rise to an Array of Different CTL Population Structures, Even When Epitopes Preserve Relative Rank Order With Respect to Immunogenicity 113
5.3.2 Immunogenicity Profile Can Have a Significant Impact on Viral Burden Because it Affects How the Rate of Spontaneous Viral Expression Differentially Manifests Itself 119
5.3.3 Comparison of Cellular-Mediated Immunity Between Individuals: What Determines an Effective CTL Response? 123
5.4 Comparison of Model Outcomes with Data 134
5.4.1 Experimental Data From Published Literature 134
5.4.2 Linking Experimental Measurements to Model Quantities 138
5.4.3 Parameter Estimation Using the Multiple Epitope Model 140
5.5 Discussion of Main Conclusions Drawn from the General Multiple Epitope Model 155
5.6 Final Remarks on the General Multiple Epitope Model 161

6 A Multi-Locus Approach With Antigenic Variation: The Multi-Locus Model 163
6.1 Motivation 164
6.2 The General Multi-Epitope Model Incorporating Antigenic Variability at Multiple Epitopes Simultaneously: m Variants in Tax, n Variants in HBZ . 165
6.2.1 Formulation of the General Multi-Locus Model 166
6.2.2 Comments About Infection by Multiple Strains and Defining Measures of Strain and CTL Fitness 169
6.2.3 The Special Case $n = 1$: Basic Analytical Insights Gained from the Multi-Locus HTLV-I System with Antigenic Variability in One Epitope 170
6.2.4 Characterisation of the Infection-Free Steady State in the General Multi-Locus Model 171
6.2.5 A Remark on Analytic Tractability of the General Multi-Locus Model 172
6.3 Co-existence of All Strains 173
6.4 The Simplest Example of a Multi-Locus System: 2 Variants in Tax, 2 Variants in HBZ 175
6.4.1 Differences in Inter-Strain Fitness Create Patterns of Viral Strain Structure 175
6.4.2 Discrete Strain Structure, Strain Dominance, and Stratification of the HTLV-I-Specific CTL Response 177

6.4.3 Differences in Effective Expansion Rates Between CTLs Determine Strain and CTL Regions . . . 183
6.4.4 Heterogeneity in the Efficiencies of Epitope-Specific CTL Responses Shapes the Regions for Ordered Strain and CTL Structure . . . 186
6.4.5 Proviral Load, Viral Composition, and Strain Dominance Are Differentially Influenced by CTLs Depending on Type of Strain Structure 188
6.5 Visualisation of Strain Structure in a Higher Order Multi-Locus System: The 3×3 Model . . . 190
6.6 Conclusions and Discussion of the Results from the Multi-Locus Model . . . 197
6.6.1 The Emergence of Ordered Strain Structure in Persistent HTLV-I Infection . . . 197
6.6.2 Why Strain Structure? . . . 198
6.6.3 Virus Composition and Diagnosis of HTLV-I-Associated Pathology . 199
6.6.4 Both Virus and Host Factors Contribute to the Establishment and Maintenance of Strain Structure . . . 201
6.6.5 Implications for Administration of Vaccines . . . 203
6.7 Final Remarks on the Multi-Locus Model . . . 204

7 Conclusions and Future Directions 206
7.1 Perspectives and Overview . . . 206
7.2 Extensions and Future Directions . . . 210
7.3 Concluding Remarks . . . 211

A Additional Derivations and Proofs 213
A.1 Derivation of the (Dimensional) Basic Reproduction Number R_0 . . . 213
A.1.1 Method I. Using the Next Generation Operator . . . 213
A.1.2 Method II. Using the Stability Condition for P_0 . . . 214
A.1.3 Method III. Using the Biological Interpretation . . . 214
A.2 Proofs for the Tax/HBZ Model (4.1) . . . 215
A.2.1 Derivation of Expression for Tax-Specific CTL Dominance in Subsection 4.3.1 . . . 215
A.2.2 Proof of Theorem 4.3 . . . 216
A.3 Proofs for the General Multiple Epitope Model (5.1) . . . 217
A.3.1 Biologically Realistic Region for the Multiple Epitope Model (5.1) . 217
A.3.2 Proof of Theorem 5.2. Part 1: Existence of Equilibria . . . 218
A.3.3 Proof of Theorem 5.2. Part 2: Stability of P_0 . . . 219

A.3.4 Proof of Theorem 5.2. Part 3: Stability of P^* 224
A.4 Proofs and Derivations for the Multi-Locus Model (6.1) in the Special Case When $n = 1$: The '$m \times 1$' Model, Corresponding to One Antigenically Variable Epitope and One Conserved Epitope 228
A.4.1 The '$m \times 1$' Model Notation and Equations 228
A.4.2 Global Properties of the Infection-Free Equilibrium for the $m \times 1$ Model 230
A.4.3 Focussing on Immunogenicity 237
A.4.3.1 Coexistence 237
A.4.3.2 The extent of Tax immunogenicity determines the relative abundances of virus strains and CTLs 240
A.4.4 Focussing on Competition Between Strains 245
A.4.4.1 The extent of immunogenicity of the advantageous strain determines co-existence or competitive exclusion of virus strains 246
A.5 Proofs for the General Multi-Locus Model (6.1) 249
A.5.1 Derivation of the Expression Relating Equal CTL Frequencies from Section 6.3 249
A.5.2 Statement and Proof of Degeneracy in the Multi-Locus Model 252

B Lemmas **255**

C Additional Figures and Tables **257**
C.1 Comparison of Exponential and Logistic Growth Profiles for the Proliferation of Actively Infected Target Cells 257
C.2 Original Tables of Experimental Data Used in the Multiple Epitope Model . 261
C.3 Additional Bar Charts and Combined Plots for Parameter Estimation Using the Multiple Epitope Model 264
C.4 Strain Structure in an Asymmetric Multi-Locus System: The 3×2 Model . 267

Bibliography **271**

List of Figures

1.1 Flowchart of the chapters in this thesis. 7

2.1 The target cell population can be categorised into three broad classes. . . . 11
2.2 A schematic representation of the two routes of viral transmission. 12
2.3 An illustration of the interaction between infected target cells and CTLs. . 15
2.4 A schematic diagram of the model by Wodarz et al. (1999). 18
2.5 A schematic diagram of the model by Asquith and Bangham (2000). 20
2.6 A schematic diagram of the model by Gómez-Acevedo et al. (2010). 21
2.7 A schematic diagram of the model by Li and Lim (2011). 22

3.1 Biological mechanism of HTLV-I infection *in vivo*. 28
3.2 Numerical simulation of the solution of model (3.1) 35
3.3 Convergence to P_R^* when $R_{0R} > 1$ 48
3.4 Piecing together transient solutions with the chronic infection equilibrium P_R^*. 52
3.5 The effect of viral activation on the proviral load. 55
3.6 A faster rate of proviral expression increases the active proportion of proviral cells. 57
3.7 Positive correlation between CTL frequency and the proviral load. 61
3.8 Negative correlation between rate of CTL-mediated lysis and the proviral load. 62
3.9 Efficient CTLs do not necessarily decrease the active proportion of proviral cells. 64
3.10 A hypothesis for determining clinical status. 66
3.11 Bifurcation diagram relating proviral load and the active proportion of infected cells. 68
3.12 Failure of HAM/TSP treatment. 69
3.13 Immune compromise leading to development of disease is possible, but unlikely. 71
3.14 A plausible mechanism for pathogenesis. 72

4.1 The effect of proviral activation on CTL-virus interactions in the Tax/HBZ model (4.1). 82
4.2 Modelling the simultaneous expression of Tax and HBZ in HTLV-I infection. 83
4.3 A more detailed look at the two phases of cell-mediated adaptive immunity. 87
4.4 Heat maps showing the strength of CTL dominance and the CTL to active virus ratio of Tax-specific CTLs in the two-epitope Tax/HBZ model (4.1). . 91
4.5 Heat map of proviral load demonstrating how immunogenicity influences the magnitude of the proviral load in different ways depending on the effective ratio between distinct CTL responses. 94
4.6 Relative immunogenicities of Tax and HBZ affect CTL-mediated control. . 96

5.1 An illustration of the key interactions in a generalised multiple epitope model of HTLV-I infection. 103
5.2 Constructing a visual representation of the properties that define an individual's anti-viral cellular immune response. 110
5.3 The structure of the CTL response for various different immunogenicity profiles within the same individual. 117
5.4 Exploring the role of CTL heterogeneity arising from differences in immunogenicity profiles for the same individual on the outcome of infection with respect to two measures of viral burden. 121
5.5 CTL properties and immunogenicity profiles for the immune responses of five different individuals. 124
5.6 CTL dominance hierarchy and its effect on the virus population across five different sample individuals. 125
5.7 CTL properties and immunogenicity profiles for the immune responses of the five individuals considered earlier, except with uniform immunogenicity profiles. 132
5.8 CTL dominance hierarchy and its effect on the virus population across the same five sample individuals considered in the previous figure, but with uniform immunogenicity profiles. 133
5.9 Bar chart of the experimental data from Asquith et al. (2005a,b) describing the viral burden for a cohort of HTLV-I seropositive individuals. 142
5.10 List plot representations of the immunogenicity profiles for the experimental data by Goon et al. (2004a). 146
5.11 Matching the multiple epitope model to experimental data allows us to derive estimates for two key parameters. 150

6.1 A multi-locus approach to modelling HTLV-I infection involves the consideration of antigenic variability at multiple epitopes simultaneously. 166
6.2 A schematic representation of the key interactions assumed in the multi-locus model. 167
6.3 In the absence of inter-strain fitness discrepancies, we observe the co-existence of all viral strains and no strain structure (NSS). 174
6.4 Heterogeneity in strain fitness creates structure in the viral population. . . 176
6.5 The degree of heterogeneity in the fitness of individual strains (i.e. $R_{0,ij}$ values) and the extent of immunogenicity towards protein variants in the Tax epitope (i.e. δ_i) play a significant role in determining the structure of both virus and CTL populations. 178
6.6 Pie chart legend for the composition of the virus population in terms of relative abundances of each individual viral strain for different degrees of inter-strain fitness discrepancies. 181
6.7 Examining virus composition for varying degrees of inter-strain fitness discrepancies. 183
6.8 Relative differences in the effective expansion rates of immune effector cells and the skew in immunogenicity towards protein variants in the Tax epitope (i.e. δ_i) determine the organisation of the virus and CTL populations in the presence of inter-strain differences in heterogeneity. 185
6.9 Heterogeneity in the efficiencies of CTL responses can change the shape of the regions of organised strain structure and CTL dominance. 187
6.10 The impact of CTL efficiency on proviral load and CTL frequency. 189
6.11 The influence of CTL efficiency on the magnitude of the proviral load and viral strain composition in the DSS and PDSS regions. 190
6.12 Heat map showing the effect of CTL efficiency on dominance of the fittest strain. 191
6.13 Pie chart legend indicating virus composition in terms of relative abundances of each individual viral strain for the higher order multi-locus model with three variants in each of Tax and HBZ. 193
6.14 Visualisation of virus composition at equilibrium in the 3×3 multi-locus model. 194
6.15 DSS and a variety of overlapping strain structures can exist in the 3×3 system. 195
6.16 DSS does not imply that the fittest strain is dominant in the 3×3 system. 196

7.1 Overview of the broad themes explored in this thesis and how each of the mathematical models relate to them. 207

A.1 Introducing antigenic variation in the Tax protein. 229

B.1 The graph of the postive definite function $f(x) = x - 1 - \log(x)$. 256

C.1 Time series plots for each of the T-cell populations under investigation comparing linear and logistic growth terms for infected target cell proliferation. 259
C.2 Time series graphs for the proviral load and active proportion of infected cells comparing linear and logistic growth terms for infected target cell proliferation. 260
C.1 Original table of experimental data from Goon et al. (2004a) (Table 1). . . 261
C.2 Experimental data from Asquith et al. (2005a) (Table 1). 262
C.3 Experimental data from Asquith et al. (2005b) (Table 1). 263
C.4 Bar charts of the experimental data from Asquith et al. (2005a,b). 264
C.5 Combined list plot representation of the immunogenicity profiles for the experimental data by Goon et al. (2004a). 265
C.6 Combined plots showing the matching of the multiple epitope model to experimental data, which allows us to derive estimates for two key parameters in each of the two HTLV-I seropositive cohorts. 266
C.7 Pie chart legend indicating virus composition in terms of relative abundances of each individual viral strain for asymmetric multi-locus model with three variants in Tax and two variants in HBZ. 267
C.8 Visualisation of virus composition at equilibrium in the asymmetric 3×2 multi-locus model. 268
C.9 DSS and strain dominance in the asymmetric 3×2 system. 269
C.10 DSS does not imply that the fittest strain is dominant in the asymmetric 3×2 system. 270

Glossary

Glossary/Table of Terminology

Table 1: Table of immunological terms, abbreviations, and acronyms that frequently appear in the text.

Term	Meaning
AC	asymptomatic carrier, an infected individual who displays no visible symptoms
Antigen	any molecule that triggers an immune response
ATL	adult T-cell leukaemia/lymphoma, an aggressive blood/lymph cancer caused by HTLV-I
CTL	$CD8^+$ cytotoxic T-lymphocyte, or a so-called 'killer T-cell', the principle branch of host immunity
CTL profile	the set of characteristics or attributes that define an individual's cellular immune response against an invading pathogen, such as the number and relative immunogenicities of epitopes that are recognised, the abundance of each CTL type (i.e. the 'immunodominance hierarchy'), and the rates of proliferation, natural death, and cell-mediated lysis of infected target cells; the 'CTL profile' encapsulates a broader scope than the 'immunogenicity profile'
CTL rank order	a list of CTL epitopes in order from least immunogenic to most immunogenic; not to be confused with 'immunodominance hierarchy', which describes the order of CTL responses in terms of their abundances
DSS	'discrete strain structure', or sometimes referred to as 'discordant strain structure'; describes the stable co-existence of a set of antigenically variable virus strains which do not share variants at any of the same epitopes
Epitope[1]	the specific part of a (viral) antigen that can be recognised by host immunity, a single antigen can have multiple distinct epitopes

Continued on next page ⟶

[1]Called a 'viral epitope', because it is a product of the virus, but can also be referred to as a 'CTL epitope', due to it being the part of the viral product which is recognised by the CTL response. Both terms are interchangeable.

Table 1 – Continued from previous page.

Term	Meaning
HAM/TSP	HTLV-I-associated myelopathy, also known as tropical spastic paraparesis, a neurological and inflammatory disease caused by HTLV-I
HBZ	HTLV-I basic leucine zipper factor, an anti-sense HTLV-I protein believed to play an important role in HTLV-I infection
HLA	human leukocyte antigen, the major histocompatibility complex in humans, aids in host immunity
HTLV-I	human T-cell lymphotropic virus type 1
Immunodominance	a phenomenon occurring during infection by a pathogen in which the bulk of the host immune response is directed towards only one or two epitopes, even though more epitopes are present
Immunodominance hierarchy	a ranked list of epitope-specific cellular immune responses according to CTL frequency (or abundance); the most abundant CTL type is said to be 'immunodominant', whilst the remaining, less prevalent, CTL types are called 'subdominant'
Immunogenicity	the extent to which an antigen or epitope provokes an immune response
Immunogenicity profile (IP)	the set or collection of relative immunogenicities for all viral epitopes recognised by a given individual's immune system; not to be confused with 'CTL profile', which describes the collective properties of the immune response
KIR	stands for 'killer-cell immunoglobulin-like receptors'; regulates the function of immune cells, usually associated with innate immunity, but recent evidence suggests that KIRs can also regulate aspects of adaptive immunity
NSS	'no strain structure'; often used to describe the situation in which all possible strains co-exist
PBMC	peripheral blood mononuclear cell, which includes any blood cell that has a round nucleus (as opposed to a lobed nucleus). For example, $CD4^+$ helper T-cells, monocytes, macrophages, B-cells, $CD8^+$ T-cells, and NK cells
PDSS	'partial discrete strain structure'; describes a stable set of strains that neither contains all possible strains nor is strictly discrete; in other words, the set of strains consists of a subset of discrete strains, but it also contains other strains that exhibit some degree of overlap in at least one of the epitopes
Proviral load (PVL)	the proportion of PBMCs that carry a copy of the HTLV-I genome, a measure of viral abundance
Provirus	integrated viral DNA
Target cell	$CD4^+$ helper T-cell, the primary target cells of HTLV-I
Tax	the immunodominant HTLV-I protein

Chapter 1

Introduction

1.1 HTLV-I Infection

Human T-cell lymphotrophic virus type I (HTLV-I) is a persistent human retrovirus that infects between 10–25 million individuals world-wide (Bangham, 2000a; Gallo, 2005; Gonçalves et al., 2010; Mortreux et al., 2003; Proietti et al., 2005). It is the causative agent of a wide range of chronic inflammatory afflictions, and has been clinically linked to two major, independent diseases: adult T-cell leukaemia/lymphoma (ATL), an aggressive cancer of the blood or lymph organs that often leads to mortality within a year from diagnosis, and HTLV-I-associated myelopathy or tropical spastic paraparesis (HAM/TSP), a slowly progressive neurological and inflammatory disease that causes gradual perivascular demyelination and axonal degeneration in the central nervous system of chronically infected patients, eventually leading to severe motor dysfunction and permanent disability (Bangham, 2000a; Gallo, 2005; Levin and Jacobson, 1997; Mortreux et al., 2003; Proietti et al., 2005). Despite significant advances over the past three decades, there still remain questions about the way in which HTLV-I-infected cells successfully evade a vigorous and chronically activated virus-specific host immune response mediated primarily by cytotoxic T-lymphocytes (CTLs) or so-called 'killer T-cells' (Asquith and Bangham, 2007, 2008; Bangham et al., 1999, 2009). There is no cure nor preventative vaccine for the acquisition of infection, and neither is there effective treatment available for HTLV-I-associated pathologies — infection is life-long (Bangham, 2000a; Kubota et al., 2007; Mosley et al., 2005; Proietti et al., 2005). Intriguingly, the vast majority (over 90%) of provirus-positive individuals are life-long asymptomatic carriers of the virus. The remaining 5–10% develop ATL or HAM/TSP; however, the exact mechanisms for pathogenesis are unknown and a precise determinant of disease outcome has not yet been established (Asquith and Bangham, 2000; Bangham and Osame, 2005; Matsuoka and Green, 2009; Mosley et al., 2005).

A central issue in current HTLV-I research is how the virus is able to persist despite the presence of strong immune pressure. We will use mathematical modelling methodology to shed light on this issue. In this thesis, we develop a framework to address specific aspects of persistent viral infection in the presence of host-mediated immunity that have not been previously considered in mathematical models of HTLV-I. Although our approach is primarily theoretical in nature, this research is principally driven by a desire to elucidate biologically and clinically relevant phenomena observed in HTLV-I. We construct a series of mathematical models designed to explore the host-virus interface which, over the length of this thesis, progressively integrate the following three biologically significant features or 'themes': (I) proviral latency and activation, (II) simultaneous viral protein expression, and lastly, (III) antigenic variation. We explore and build upon these 'themes'; that is to say, as we progress through more and more complicated models, we are mindful of what we have learnt in the previous models.

Theme I: Proviral Latency and Activation

A traditional viewpoint of the survival strategy of HTLV-I has been one of complete viral latency — infected cells carry transcriptionally silent proviruses and evade destruction because they are 'invisible' to immune-mediated surveillance (Asquith et al., 2000; Richardson et al., 1997). However, such a scenario cannot adequately explain the clinical observation that blood samples taken from HTLV-I seropositive individuals typically show high frequencies of chronically activated HTLV-I-specific cytotoxic T-lymphocytes (CTLs), or so-called 'killer T-cells', as the maintenance of such a vigorous immune response requires continuous stimulation by proviral cells that are actively expressing viral antigens (Asquith et al., 2005a; Bangham et al., 2009; Daenke et al., 1996; Niewiesk et al., 1995; Parker et al., 1994). From this observation, Asquith et al. (2000) raised the natural question: 'Is HTLV-I really silent?' Several years later, Asquith et al. (2007) demonstrated experimentally that HTLV-I is, in fact, not a completely latent infection and, moreover, the authors have proposed an experimental hypothesis that HTLV-I persists by achieving a highly dynamic balance between viral activation and transcriptional latency (Asquith et al., 2007; Asquith and Bangham, 2008). This balance is one of the underlying features which we incorporate in all of the models developed in this thesis, and we examine its implications in detail, especially in the formulation and characterisation of our so-called 'baseline model' (3.1), which captures the fundamental dynamics of persistent HTLV-I infection.

Theme II: Simultaneous Viral Expression

In the next step of our work, we will focus on illuminating the role of simultaneous expression of multiple distinct viral proteins. With respect to HTLV-I, two such key viral proteins, namely Tax and HBZ, are believed to determine the outcome of the infection (Boxus and Willems, 2009; MacNamara et al., 2010). Although the functions of the immunodominant viral protein Tax have been studied extensively, until recently relatively little has been known about the impact of HBZ. It is now believed that expression of the much less immunogenic viral protein HBZ could play a crucial role in determining the outcome of infection, including the level of detriment due to the virus as well as the risk of developing disease (MacNamara et al., 2010; Matsuoka and Green, 2009; Saito et al., 2009). In addition, experimental research has shown that activated HTLV-I-infected cells can down-regulate the expression of Tax whilst maintaining or even up-regulating the expression of HBZ (Boxus and Willems, 2009; MacNamara et al., 2010; Matsuoka and Green, 2009).

In fact, it is known that a range of viral proteins, roughly 12, can be simultaneously expressed by an activated HTLV-I-infected target cell (Boxus and Willems, 2009; Goon et al., 2004a). Whereas in the baseline model (3.1) we have represented the anti-HTLV-I cellular immune response by a single homogenous pool of CTLs, the consideration of multiple epitopes requires a more complex immune system dynamic that incorporates a range of distinct types of CTL responses, each of which can only target a specific viral antigen. We explore the resulting heterogeneity in the pool of anti-HTLV-I immune cells.

Theme III: Antigenic Variability

From a more general viewpoint, it is known that each of the two key viral proteins, Tax and HBZ, display variation in their genomic sequences (Gaudray et al., 2002; Kubota et al., 2007; Niewiesk et al., 1994), and such variation can arise due to the strong, positive selection forces exerted by antigen-specific CTLs (Furukawa et al., 2001; Niewiesk and Bangham, 1996). A further extension to our HTLV-I models that takes into account the interactions between an HTLV-I provirus displaying multiple variable epitopes and variant-specific CTLs better reflects the antigenically diverse nature of HTLV-I and could therefore help resolve several unknown issues in the way in which HTLV-I-infected target cells navigate the complex within-host environment and persist despite the presence of host immunity. In this part of our investigation, we first examine the basic principles accompanying antigenic variability in only one of the two epitopes before moving to the scenario where variability exists in both epitopes.

1.2 Mathematical Methodology

Undoubtedly, one of the greatest advantages to modelling the HTLV-I system using a mathematical, rather than a verbal, approach is that we can make use of mathematical techniques to rigorously and unambiguously characterise model behaviour, allowing for a systematic way to draw inferences from the model results. To better understand the dynamics of our mathematical models, we utilise standard mathematical and computational methods, including non-dimensionalisation, stability, asymptotic, and bifurcation analyses, to investigate the various systems analytically and numerically. In each of our approaches, we can identify a threshold parameter, the basic reproduction number for viral infection R_0 (or in later chapters when we consider antigenic variability, the basic reproduction number for a given virus strain), which is a measure of viral fitness or the ability of the virus to propagate within the host. For the mathematical models that consider a single strain (i.e. in the absence of antigenic variability), we are able to prove that viral fitness in terms of the basic reproduction number completely characterises the global behaviour of solutions to the model: if $R_0 < 1$, the infection is cleared; if $R_0 > 1$, the infection is chronic. The global stability of the respective equilibria in each of the two cases for R_0 has been shown by constructing appropriate Lyapunov functions.

Establishing the global dynamics of the mathematical models not only allows us to fully investigate the finer aspects of these models without worrying about unexpected behaviour, it also provides us with a clear understanding of the type of behaviour that one might expect in later models where antigenic variability is taken into consideration, for which the global dynamical behaviour cannot always be characterised analytically. Because the mathematical models have been built on a solid foundation and motivated by immunological processes, our approach therefore places us in a position where we can ask biologically meaningful questions and then use our models to try and answer them. The results of our investigation yield biologically significant conclusions that either link closely to clinical observations or may be experimentally testable, and furthermore suggest novel hypotheses for the classification and control of virus-associated pathology.

1.3 Objectives and Overview

The overall objective of this thesis is to develop a consistent theoretical framework that can help shed light on specific, biologically relevant questions that are of interest to experimentalists and theoretical immunologists trying to understand the complicated host-pathogen dynamics of chronic viral infections such as HTLV-I. We will investigate these questions using a mathematical modelling approach. Mathematical modelling can help us break apart

the complex mechanisms of HTLV-I persistence and identify the underlying principles that govern successful viral propagation in the presence of host immunity, providing insights on the evolution of HTLV-I infection and virus-associated pathogenesis. Understanding these interactions is a crucial step on the road to developing effective ways to disrupt the virus life-cycle and may help identify promising new treatment strategies to reduce the severity of HTLV-I infection and minimise the detriment due to associated disease. Although the issues here have been investigated in the context of persistent HTLV-I infection, the framework that we have established can be applied more generally to a range of pathogenic invasions. In the following, we outline the structure of this thesis. A flowchart of the various chapters is shown in Figure 1.1.

In the present Introduction chapter (Chapter 1), we have set out a broad motivation behind our study of within-host HTLV-I infection, outlined three overlapping themes which we will explore, briefly touched upon the mathematical methodology that will be used, and now provide a general overview of the contents of this thesis. In Chapter 2, we present a thorough immunological survey of the virus, focussing in particular on the properties of infected cells and their interactions with host immunity. We also review some of the key mathematical modelling work that has previously been done on HTLV-I. The biological details described here serve as an important basis for how we construct our mathematical models in order to accurately capture the host-virus dynamic.

Throughout Chapters 3 to 6, we develop a series of four mathematical models of HTLV-I infection which incorporate an increasing level of biology detail covering the three themes mentioned above. In Chapter 3, we present our 'baseline model', which is represented by a four-dimensional non-linear system of ordinary differential equations describing the dynamic interactions among viral expression, infected target cell activation, and the HTLV-I-specific CTL response. The baseline model focusses on elucidating Theme I and, in particular, the roles of viral expression and host immunity during the chronic phase of HTLV-I infection to gain some insights on viral persistence and pathogenesis. This mathematical model serves as a foundation for the models in subsequent chapters. In Chapter 4, we begin our foray into Theme II by considering the simultaneous expression of two distinct epitopes, leading to the development of the basic two-epitope 'Tax/HBZ model' of HTLV-I infection. The consideration of multiple epitopes in HTLV-I infection naturally raises further questions about HTLV-I immunology and allows us to better capture the essential nature of HTLV-I infection and persistence. Following this, there are two different directions which we can take. Along one direction, in Chapter 5, we continue developing the ideas of multiple viral expression in more detail by generalising to the situation with an arbitrary number of distinct epitopes, which we called the 'multiple epitope model'. Along the other direction, we

explore Theme III by incorporating antigenic variation in the two-epitope model. Namely, in Chapter 6, we consider antigenic variability in the two epitopes, which results in the development of the first 'multi-locus model' for within-host HTLV-I infection.

In Chapter 7, we wrap up the body of the thesis by briefly summarising the main conclusions of our work and listing a number of directions and model extensions that can be explored in more detail in the future. Lastly, we finish off the thesis with several concluding remarks.

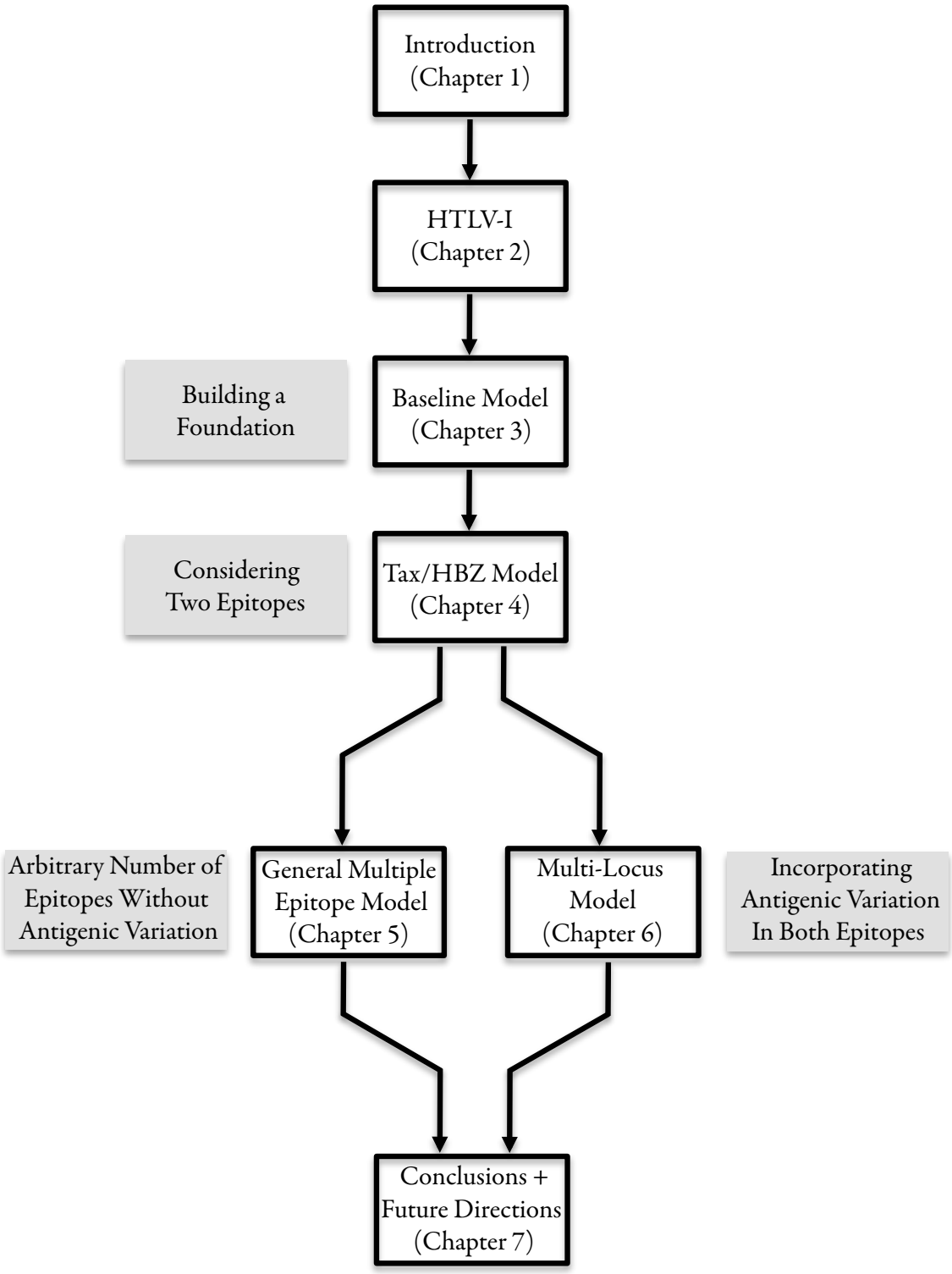


Figure 1.1: Flowchart of the chapters in this thesis showing how each of the chapters relate to one another.

Chapter 2

HTLV-I

In this chapter we present a thorough epidemiological, immunological, and mathematical survey of HTLV-I. We begin by briefly describing the epidemiology of HTLV-I infection and its associated diseases. As our focus is on elucidating the within-host dynamics of HTLV-I infection and persistence, we proceed in the following sections by discussing essential properties of virus-infected cells, including the methods of viral transmission and the roles of different HTLV-I proteins in the course of the infection. Next, we describe the properties of the HTLV-I-specific immune response, which is believed to play a significant role in the outcome of the infection. Lastly, we highlight several key mathematical models that have been previously considered and motivate the formulation of our own baseline mathematical model in the following chapter. A table of biological terms, abbreviations, and acronyms that commonly appear in this chapter and the remainder of the thesis can be found in the Glossary (Table 1).

2.1 Epidemiology

Human T-lymphotrophic virus type I (HTLV-I) was the first discovered human retrovirus, identified independently in Japan and the United States in the late 1970's to early 1980's, and it has been estimated that between 10–25 million individuals world-wide are infected with HTLV-I (Bangham, 2000a; Gallo, 2005; Gonçalves et al., 2010; Mortreux et al., 2003; Proietti et al., 2005). Since its discovery, the geographic distribution of HTLV-I has been studied extensively, with southern Japan, the Carribean, Africa, Central and South America being identified as regions where the infection is endemic (Bangham, 2000a; Gonçalves et al., 2010; Proietti et al., 2005; Tomaru et al., 2005). The transfer of infected cells either through breast milk, semen, or blood is necessary for infection to occur horizontally between individuals, and infection can also be transmitted vertically from mother to child (Bangham, 2000a; Bangham et al., 2009; Proietti et al., 2005).

The vast majority (over 90%) of HTLV-I-infected individuals are asymptomatic carriers (ACs) of the virus and do not display symptoms. The remaining 5%–10% develop a form of HTLV-I-associated pathology. HTLV-I is the aetiological agent of two major, clinically independent diseases which usually develop after a long asymptomatic phase, and it is still not known what triggers disease manifestation (Asquith and Bangham, 2007; Mortreux et al., 2003). The first major pathology is adult T-cell leukaemia/lymphoma (ATL), an aggressive T-cell malignancy characterised by the emergence of atypical lymphoid cells with multi-lobulated nuclei known as 'flower cells' (Boxus and Willems, 2009; Gonçalves et al., 2010; Mortreux et al., 2003). ATL encompasses a spectrum of the disease that is divided into four clinical subtypes, namely smouldering, chronic, lymphoma, and acute (Boxus and Willems, 2009; Gonçalves et al., 2010; Mortreux et al., 2003; Proietti et al., 2005). Patients diagnosed with the most aggressive forms, the lymphoma and acute subtypes, comprise the major proportion of all ATL cases, up to 75% or higher, and typically die within 12 months, while the relatively small proportion of those diagnosed with the two remaining subtypes have a longer life expectancy, up to 24 months for the chronic type and up to five years or more for the smouldering type (Gonçalves et al., 2010; Proietti et al., 2005). The second major clinical disease is HTLV-I-associated myelopathy, also known as tropical spastic paraparesis (HAM/TSP), a slowly progressive, debilitating neurological and inflammatory disorder caused by perivascular de-myelination and axonal degeneration (Bangham et al., 1999; Levin and Jacobson, 1997). HAM/TSP is characterised by spastic gait, sensory dysfunction, hyper-reflexia, and muscle weakness, including degeneracy of the sphincter muscles which often affects the autonomic functions of the bladder and bowel (Gonçalves et al., 2010; Proietti et al., 2005). HTLV-I has also been clinically linked to a wide range of other chronic and acute diseases, including arthritis, uveitis, infective dermatitis, polymyositis, Sjögren's syndrome, and pulmonary infiltrative pneumonitis (Parker et al., 1994; Proietti et al., 2005; Tomaru et al., 2005), and is associated with an increased risk of opportunistic infections — pathogenic invasions that cannot manifest into disease in a healthy individual, but thrive in the presence of a weakened immune system — such as *Strongyloides stercoralis*, *Staphylococcus aureus*, and *Mycobacterium tuberculosis* (Asquith and Bangham, 2007; Proietti et al., 2005). In our study, we will consider HTLV-I infection principally as a precursor to the acquisition of chronic inflammatory disease, most notably HAM/TSP. Although the dynamics of T-cell malignancy are fundamentally different from non-malignant disease and alternative realisations are needed to describe leukaemogenesis, some of our results may also have biological implications that are related to the events leading up to the development of ATL.

2.2 Classification and Viral Dynamics

HTLV-I belongs to the family *Retroviridae* consisting of enveloped RNA viruses that create DNA from their original RNA genome, which is then integrated into the host cell's genetic code. The family of retroviruses includes the well-known human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), and bovine leukaemia virus (BLV). More specifically, HTLV-I is a complex, exogenous, oncogenic RNA retrovirus classified in the genus *Deltaretrovirus* of the subfamily *Orthoretrovirinae* (Bangham, 2003). The HTLV-I genome, which is 9 kb in length, is highly genetically stable with few sequence variations (Asquith and Bangham, 2008; Bangham et al., 1999; Kubota et al., 2007; Mortreux et al., 2003), and encodes the usual retroviral genes such as *gag, pol, pro* and *env*, the viral enzymes polymerase, integrase, protease and reverse transcriptase (Bangham, 2000b), accessory proteins p12, p13 and p30 (Boxus and Willems, 2009), along with sequences for regulatory proteins including Tax, Rex, Rof, Tof, and the recently identified HTLV-I basic leucine zipper factor (HBZ) (Boxus and Willems, 2009; Gaudray et al., 2002; Mosley et al., 2005). HTLV-I preferentially infects $CD4^+$ helper T-cells *in vivo*, and although significant populations of $CD8^+$ T-cells and dendritic cells may also be infected (Bangham, 2003; Boxus and Willems, 2009; Höllsberg, 1999), productive viral replication is thought to only occur in the $CD4^+$ helper T-cell population (Bangham, 2000a; Bangham et al., 1999). We will model the dynamics of HTLV-I infection in $CD4^+$ helper T-cells as the target cells.

As with other RNA retroviruses, HTLV-I uses the enzymes *reverse transcriptase* and *integrase*, respectively, to transcribe a double-stranded DNA copy of its viral genome, and to integrate this viral DNA into the host cell genome, thus allowing the virus to take advantage of the T-cell machinery and propagate the infection (Mosley et al., 2005). The integrated viral DNA is called a *provirus*, and the proportion of peripheral blood mononuclear cells (PBMCs) that carry a provirus is known as the *proviral load*. Experimental evidence has shown that cell-free virions are typically undetectable in the peripheral blood and are inefficient at infecting $CD4^+$ helper T-cells, the primary targets of HTLV-I (Igakura et al., 2003). The cell-to-cell spread of HTLV-I does not require the formation of enveloped extracellular virions, thus the proviral load is an accurate measure of the viral burden in an infected individual (Asquith and Bangham, 2008; Mosley et al., 2005). A characteristic feature of chronic HTLV-I infection is the establishment of high proviral loads; the proportion of proviral cells in the periphery typically falls in the range of 0.1%–10% PBMCs, but may approach 5% of PBMCs in an AC, and may be as high as 40% PBMCs or up to 50%–70% of $CD4^+$ helper T-cells in HAM/TSP patients (Asquith and Bangham, 2007; Asquith et al., 2000; Bangham, 2000a; Bangham and Osame, 2005; Mortreux et al., 2003; Mosley et al.,

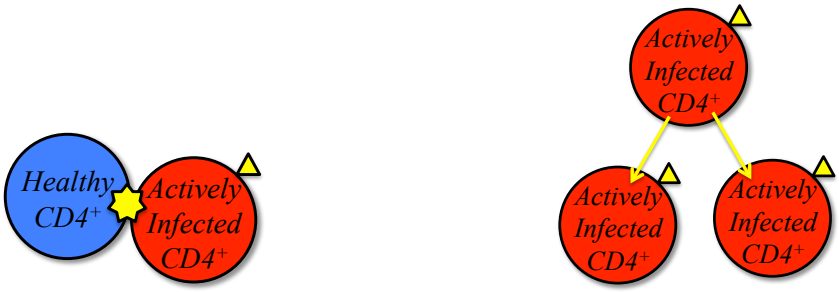
2005). In contrast, during the later stages of acquired immunodeficiency syndrome (AIDS), the proviral load in HIV-infected individuals rarely exceeds 3.3% PBMCs (Mortreux et al., 2003).

HTLV-I proviral cells may be separated into two broad classes depending on their state of activation, which is accompanied by the suppression or expression of viral antigens: (i) latently infected target cells are resting $CD4^+$ helper T-cells that contain a provirus but do not display viral proteins, and (ii) actively infected target cells are activated provirus-harbouring $CD4^+$ helper T-cells that do display viral proteins. While latently infected cells that hide viral antigens are transcriptionally silent and do not make copies of the provirus, actively infected cells that express viral proteins undergo persistent replication of the provirus. Experimental work by Richardson et al. (1997) has demonstrated that viral latency is frequent and that such latently infected proviral cells do not undergo selective, spontaneous proliferation unlike their actively infected counterparts. In other words, there is a correlative link between viral gene expression and replication of the provirus. Figure 2.1 shows a cartoon illustration of the three types of target cells: healthy, latently infected, and actively infected. These cartoon representations will appear throughout the thesis.

Figure 2.1: The target cell population can be categorised into three broad classes. The circles represent $CD4^+$ helper T-cells, the primary target cells of HTLV-I infection. Healthy cells (blue) do not contain a provirus. Latent proviral cells (green) hide viral protein expression (represented by yellow triangles). Activated proviral cells (red) display viral antigens.

There are two routes of viral transmission of $CD4^+$ helper T-cells: 'infectious' or horizontal, and 'mitotic' or vertical. These are illustrated schematically in Figure 2.2. Infectious transmission occurs through direct cell-to-cell contact between an activated proviral cell and a healthy cell. Igakura et al. (2003) showed experimentally that a reorientation of the infected cell's cytoskeleton results in the formation of a highly organised structure known as a 'virological synapse' which creates a tight junction across which enveloped HTLV-I virions are efficiently transferred to the target cell without escaping into the periphery. Once entry into the host cell cytoplasm is complete, the HTLV-I virion unpackages itself, along with viral enzymes such as reverse transcriptase and integrase, and proceeds to incorporate itself into the host cell genome. Mitotic transmission occurs when an infected parent cell undergoes cellular division, creating two identical daughter cells that both carry an integrated copy of the viral DNA. Experiments have shown that the vast majority of

an infected individual's proviral load consists of large numbers of clonally expanded copies of a relatively small number of different T-cell clones (Mortreux et al., 2003). Because of this observation, the relative lack of variability in the HTLV-I genetic sequence within an infected individual over time, and since reverse transcription is known to be an error-prone process, it is reasonable to infer that during chronic HTLV-I infection, the proviral set point is maintained primarily by the mitotic route of infection (Mortreux et al., 2003). However, the existence of distinct T-cell clones in an infected individual, along with evidence of ongoing viral replication *in vivo*, suggests that infectious transmission is also a substantial contributing factor to maintenance of the proviral load and its impact on the outcome of infection should not be ignored (Bangham, 2003; Bangham et al., 1999; Mortreux et al., 2003; Niewiesk et al., 1995). Indeed, conventional experimental techniques have previously shown that T-cell clones number at least in the hundreds (Mortreux et al., 2003). Recently, the invention of more precise measurement techniques has led to the discovery that such an estimate is far too modest and that identification of infected T-cell clones could potentially be on the order of thousands or ten thousands, or even more (Gillet et al., 2011).



(a) Infectious transmisson.

(b) Mitotic transmission.

Figure 2.2: A schematic representation of the two routes of viral transmission. Healthy target cells and actively infected target cells are drawn as blue and red circles, respectively.

2.3 HTLV-I Proteins

There are approximately 12 HTLV-I proteins in total, including the structural proteins Gag and Env, as well as Pol and Pro, which are common to all retroviruses. In addition, the HTLV-I genome encodes several non-structural regulatory proteins Tax, Rex, Rof, Tof and HBZ, as well as the accessory proteins p12, p13 and p30 (Boxus and Willems, 2009). Among all the HTLV-I proteins, Tax and HBZ appear to play crucial roles in viral persistence and in the progression of disease (Boxus and Willems, 2009; MacNamara et al., 2010; Matsuoka and Green, 2009; Saito et al., 2009).

The regulatory protein Tax (p40) is one of the first viral proteins to be expressed (Bangham et al., 1999). Tax is highly immunodominant and promotes viral replication through its transcriptional transactivator function (Bangham et al., 1999). It is responsible for driving the expression of both host and viral genes, and induces infected T-cell proliferation via up-regulation of cellular genes involved in mitosis, down-regulation of cell-cycle checkpoints, and reducing the threshold of T-cell activation (Asquith and Bangham, 2008; Asquith et al., 2007; Kubota et al., 2007; Mortreux et al., 2003; Niewiesk et al., 1995; Wattel et al., 1996). Tax has also been shown to inhibit DNA repair and generate DNA abnormalities, leading to genetic instability and subsequent transformation of infected target cells (Boxus and Willems, 2009; Matsuoka, 2005; Mortreux et al., 2003). Despite its oncogenic role in disrupting the stability of the host cell genome, Tax is itself functionally constrained, and even minor mutations can impair its transactivation abilities (Kubota et al., 2007; Niewiesk et al., 1995).

Recently, the regulatory protein HTLV-I basic leucine zipper factor (HBZ) has been identified as having a significant impact on proviral load and the risk of developing HTLV-I-associated pathologies, both HAM/TSP and ATL (MacNamara et al., 2010; Matsuoka and Green, 2009; Saito et al., 2009). Unlike the other HTLV-I proteins, HBZ is encoded by the complementary (anti-sense) strand of the HTLV-I genome (Gaudray et al., 2002; Matsuoka and Green, 2009; Saito et al., 2009). It has been observed that HBZ expression promotes infected T-cell proliferation while suppressing Tax-mediated viral transactivation through the $5'$ long terminal repeat (LTR) (Boxus and Willems, 2009; Matsuoka and Green, 2009; Saito et al., 2009). The expression of HBZ and its connection with the immunodominant viral Tax protein is an important factor that we will consider using mathematical modelling.

The remaining regulatory and accessory proteins also play significant parts in establishing and maintaining HTLV-I infection. Rex (p27), whose coding sequence substantially overlaps that of Tax, is a post-transcriptional regulatory protein that mediates the transport and splicing of viral mRNA (Bangham, 2000a,b; Boxus and Willems, 2009; Smith et al., 1997). The HTLV-I regulatory proteins Rof and Tof are believed to play an important role in the establishment of chronic HTLV-I infection (Pique et al., 2000). The accessory proteins are involved in various processes and have been the subject of previous investigation (Johnson et al., 2001). For example, p30 appears to modulate the expression and transcription of viral and cellular genes and, along with its truncated form p13, aids in the maintenance of high proviral loads *in vivo* (Johnson et al., 2001; Pique et al., 2000). The p12 protein displays multiple functions, including regulation of calcium-dependent signalling and up-regulation of transcriptional activity, and has been shown to cause intracellular degradation of MHC Class I molecules, preventing the presentation of HTLV-I antigens and possibly

reducing recognition by host immune responses (Bangham et al., 2009; Johnson et al., 2001; Pique et al., 2000).

2.4 HTLV-I-Specific Host Immunity

As with most pathogenic invasions, HTLV-I infection induces innate and adaptive immune responses, including recruitment and activation of macrophages, dendritic cells, B-cells, and T-cells. Strong adaptive immune responses to HTLV-I are established within 7–10 days, and consist mainly of the humoral or antibody response, and the cellular or cytotoxic T-cell response. Recent research has also shed light on the role of the $CD4^+$ helper T-cell response during infection (Bangham, 2003).

HTLV-I-infected individuals have appreciable levels of HTLV-I-specific antibodies, and their detection is the basis of diagnosis of HTLV-I infection (Bangham, 2000a; Mosley et al., 2005). Antibodies against dominant epitopes on the viral protein Tax are the most abundant, although antibodies targeting Gag and Env are also present (Bangham, 2000a,b). However, the role of the humoral response in controlling the proviral load or preventing the emergence of disease is unclear (Bangham, 2000b, 2003; Mosley et al., 2005). In addition, the antibody titre in serum against HTLV-I antigens is often remarkably high and displays a positive correlation with proviral load (Bangham, 2000b, 2003; Mosley et al., 2005), suggesting that anti-HTLV-I antibodies are at best ineffective at limiting viral replication, and may even play a part in the pathogenesis of HAM/TSP.

In contrast, cellular immunity, which is mediated primarily by the expansion and lytic effects of HTLV-I-specific $CD8^+$ cytotoxic T-lymphocytes (henceforth referred to as 'CTLs'), also known as 'killer T-cells', is hypothesised to be a major determinant of proviral load and disease outcome (Asquith and Bangham, 2000; Asquith et al., 2005a; Bangham and Osame, 2005). Chronically activated anti-HTLV-I CTLs are typically abundant in the peripheral blood of infected hosts, both ACs and HAM/TSP, suggesting persistent production and expression of HTLV-I antigens *in vivo* (Asquith et al., 2005a; Bangham et al., 2009; Daenke et al., 1996).

HTLV-I-specific CTLs eliminate activated infected target cells, and the proliferation of these anti-HTLV-I CTLs is stimulated in response to expressed viral antigens (Bangham, 2003). Latently infected target cells that do not express viral proteins evade immune surveillance and are safe from CTL-mediated lysis. These interactions are illustrated in Figure 2.3.

The immunodominant viral protein Tax is the primary antigen recognised by circulating anti-HTLV-I CTLs (Bangham, 2000a; Goon et al., 2004a; Mosley et al., 2005; Parker et al., 1994). Blood samples taken from most HTLV-I-infected individuals consistently show that

(a) CTLs and active proviral cells.

(b) CTLs and latent proviral cells.

Figure 2.3: An illustration of the interaction between infected target cells and HTLV-I-specific CTLs. (a) CTLs (grey rectangles) proliferate in response to expressed antigen and eliminate actively infected target cells (red circles). (b) Latently infected cells (green circles) cannot be seen by the immune response and therefore evade CTL-mediated lysis.

these Tax-specific CTLs occur at high frequency and are chronically activated (Asquith and Bangham, 2008; Bangham, 2003; Parker et al., 1994), suggesting that they have recently been exposed to antigen *in vivo*, and implying continuous expression of Tax in the infected target cell population (Niewiesk et al., 1995; Parker et al., 1994). The presence of CTLs specific to other HTLV-I regulatory proteins including Tof, Rof, and the minus-strand gene protein HBZ have also been detected (Bangham, 2003; MacNamara et al., 2010; Pique et al., 2000). Anti-HTLV-I CTLs can also target the usual retroviral antigens such as Gag, Pol and Env (Bangham, 2003). Interestingly, the regulatory protein Rex, whose gene overlaps with Tax, does not elicit a strong CTL response, and it is not clear why this would be the case (Daenke et al., 1996; Bangham, 2003).

As $CD4^+$ helper T-cells are the main host cells of infection, the effector functions of the $CD4^+$ T-cell response to HTLV-I has been difficult to study (Bangham, 2003). Nonetheless, it has been observed that $CD4^+$ helper T-cells specific to HTLV-I are more frequently infected than $CD4^+$ T-cells with different antigen specificities (Bangham, 2003; Goon et al., 2004b; Mosley et al., 2005).

2.5 Pathogenesis of HTLV-I-Associated Disease

Current hypotheses of HTLV-I-associated disease manifestation suggest that the magnitude of an infected individual's proviral load is directly correlated with pathology; however, the ranges of the proviral loads in ACs and HAM/TSP patients are broad and overlapping, and it is clear that a high proviral load is neither necessary nor sufficient to cause disease (Asquith and Bangham, 2007; Asquith et al., 2000; Mosley et al., 2005). Indeed, there exist ACs that display a high proviral load (characterised as greater than 3% PBMCs in-

fected) along with HAM/TSP patients that exhibit a low proviral load (characterised as fewer than 1% PBMCs infected) (Asquith and Bangham, 2007; Asquith et al., 2005b; Bangham et al., 1999; Mosley and Bangham, 2009; Nagai et al., 1998). Still, it has been observed that infected individuals with a proviral load greater than 1% PBMCs are more likely to have HAM/TSP than those whose proviral load is below 1% PBMCs, in which there is no apparent association between viral burden and the risk of disease (Bangham et al., 1999; Nagai et al., 1998). Within an infected individual, the proviral load appears to remain stable over time about a viral 'set point', which fluctuates no more than five-fold; this is contrasted with the between-individual variation, which may differ by up to four orders of magnitude, from 0.001%–40% PBMCs (Asquith and Bangham, 2008; Asquith et al., 2005a; Bangham, 2000a, 2003; Mosley et al., 2005). Since HTLV-I exhibits little variability in its genetic sequence either within the same host or between different hosts, it has been suggested that these discrepancies are likely to be due to differences in the hosts rather than in the virus, and may depend on intrinsic factors such as the human leukocyte antigen (HLA) class restriction in the host genotype (Bangham, 2000b; Bangham et al., 2009; Jeffery et al., 2000, 1999).

Cellular immunity is believed to play a role in the pathogenesis of HTLV-I-associated disease, although there is conflicting evidence as to whether CTL activity is beneficial or detrimental to the host (Asquith and Bangham, 2000; Asquith et al., 2005a; Bangham, 2003; Mosley et al., 2005; Parker et al., 1994). On the one hand, there is evidence that HTLV-I-specific CTLs control the proviral load and confer protection against the development of associated disease (Asquith and Bangham, 2000; Bangham et al., 1999; Jeffery et al., 1999). Indeed, an immunogenetics study conducted in southern Japan has shown that certain HLA class I-restricted CTL responses are significantly associated with a reduced proviral load as well as a lower prevalence of HAM/TSP (Jeffery et al., 1999, 2000). On the other hand, it has been suggested that activated HTLV-I-specific CTLs migrate through the blood-brain barrier and accumulate in the central nervous system (CNS), where they cause inflammation and damage leading up to the development of HAM/TSP (Bangham, 2000a; Biddison et al., 1997; Jacobson, 2002; Mosley et al., 2005; Mosley and Bangham, 2009). Indeed, high frequencies of HTLV-I-specific $CD8^+$ CTLs have been found in white matter lesions and in the cerebro-spinal fluid (CSF) of HAM/TSP patients (Bangham, 2003; Levin and Jacobson, 1997). The two seemingly opposing hypotheses are in fact not mutually exclusive, as the CTL response depends on a complex system of positive and negative controls in its interaction with pathogenic invasions which determines whether its effect is beneficial (e.g. clearance of pathogen) or detrimental (e.g. auto-immune disorders) (Bangham, 2003; Bangham et al., 1999).

2.6 Simultaneous Expression of Multiple Viral Proteins: Focussing on the Roles of Tax and HBZ

Traditionally, it has been argued that Tax, which elicits the most intense HTLV-I-specific CTL response, is the biggest determinant of proviral load and disease status (AC or HAM/TSP). Indeed, it has been observed that asymptomatic carriers, who generally carry a lower proviral load, exert a stronger positive selection on the *tax* gene than do HAM/TSP patients (Bangham, 2003; Kubota et al., 2007; Niewiesk et al., 1994). However, MacNamara et al. (2010) have recently demonstrated that strong CTL selection in the context of HLA class I for the less immunogenic viral protein HBZ, rather than Tax, is associated with a lower proviral load and asymptomatic carriage.

In contrast with Tax, which stimulates HTLV-I expression including that of HBZ, the role of HBZ is to down-regulate Tax-mediated viral transactivation, thereby inhibiting the expression of other HTLV-I genes (Boxus and Willems, 2009). Indeed, the anti-sense HTLV-I protein HBZ may be expressed independently of other HTLV-I antigens (MacNamara et al., 2010; Matsuoka and Green, 2009). In addition, while Tax is immunodominant and elicits a vigorous cellular immune response, CTLs targeting HBZ have been shown to be much less frequent (Suemori et al., 2009). In terms of strategies for viral persistence, these observations suggest a mechanism by which HTLV-I-infected target cells can evade destruction by host immune responses against the immunodominant protein Tax whilst benefitting from rapid, selective growth through the function of the alternative HBZ transactivator (Suemori et al., 2009).

2.7 Previous Modelling Work

Previous within-host mathematical models of HTLV-I form a basis upon which we may expand. One fundamental aspect of the dynamics of HTLV-I infection is the interaction between proviral cells and virus-specific adaptive immune responses, which are mediated primarily by $CD8^+$ CTLs. In this section, we briefly highlight some previous work in mathematical modelling of HTLV-I infection that is important to the formulation of our baseline model.

2.7.1 Wodarz, Nowak, and Bangham (1999)

Wodarz et al. (1999) have developed a mathematical model for HTLV-I infection that takes into account both horizontal and vertical transmission of the provirus in addition to an explicit consideration of the Tax-specific CTL response. The target cell population of $CD4^+$ helper T-cells is split into healthy and proviral groups, which are denoted by $x(t)$ and $y(t)$,

respectively. The pool of anti-HTLV-I CD8$^+$ CTLs is denoted by $z(t)$. The production of new, uninfected target cells from the thymus is assumed to be at a constant rate λ. Healthy and infected target cells undergo density-dependent proliferation with respective rate constants r and s. The rate r represents homeostatic proliferation of the CD4$^+$ helper T-cell population while the rate s represents mitotic transmission of the provirus in infected CD4$^+$ T-cells. Target cell proliferation is assumed to follow logistic growth behaviour, with the total number of target cells, including both healthy and infected, bounded above by a carrying capacity, k. Healthy cells that come into direct contact with infected cells may themselves become infected via the infectious route of viral transmission at a rate β, and this interaction is assumed to have a bi-linear incidence form. The proliferation of anti-HTLV-I CTLs occurs at a rate c and at high densities is assumed to be proportional simply to the number of infected target cells. Specifically, the density of CTLs saturates according to $z(t)/(z(t)+1)$. CTLs are assumed to encounter and lyse infected cells at a bi-linear incidence rate p. Lastly, the natural death rates of healthy and infected CD4$^+$ helper T-cells are $dx(t)$ and $ay(t)$, respectively, and that of CD8$^+$ CTLs is $bz(t)$. A schematic diagram of the interactions is shown in Figure 2.4.

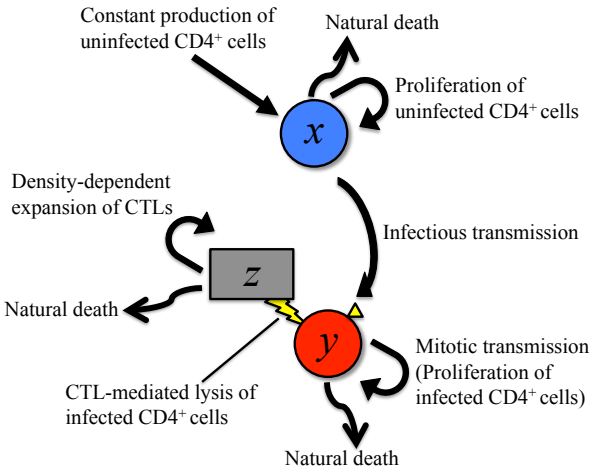


Figure 2.4: A schematic diagram of the model by Wodarz et al. (1999) that considers the interactions among healthy target cells, x, infected target cells, y, and virus-specific CTLs, z. The effects of target cell latency are not considered.

Let $'$ denote the derivative with respect to time. The above assumptions lead to the following system of equations:

$$\begin{aligned} x' &= (\lambda + rx)\left(1 - \frac{x+y}{k}\right) - dx - \beta xy \\ y' &= \beta xy + sy\left(1 - \frac{x+y}{k}\right) - ay - pyz \\ z' &= \frac{cyz}{z+1} - bz. \end{aligned} \tag{2.1}$$

Although analytical techniques were not utilised, Wodarz et al. (1999) performed extensive numerical investigations on their model, using non-negative initial conditions, to study the contributions of infectious and mitotic transmission at equilibrium. The authors concluded that although the high proviral load is maintained principally by vertical or mitotic spread, the effects of horizontal or infectious spread can also be a significant contributing factor. Their results demonstrated the proof of principle that the relative sequence invariance of the HTLV-I genome does not necessarily imply that HTLV-I replication via infectious transmission is low.

This model illustrates the importance of incorporating the role of HTLV-I-specific immune responses to fully understand the interactions that occur during chronic infection. However, this model does not address the issue of how HTLV-I-infected individuals, both ACs and HAM/TSP patients, often accumulate high proviral loads despite strong positive selection by anti-HTLV-I CTLs *in vivo* (Asquith et al., 2005a; Bangham and Osame, 2005).

2.7.2 Asquith and Bangham (2000)

Asquith and Bangham (2000) have developed a mathematical model that describes the interaction between infected target cells and the anti-HTLV-I CTL response. A notable feature of this model that differs from other models is the explicit incorporation of cytokine production by CTLs. Cytokines are a class of proteins produced by various cells that affect the behaviour of other cells (Parham, 2005). In the model, the authors consider the inflammatory cytokine, interferon-γ (IFNγ), which plays a critical role in innate and adaptive immunity against pathogenic invasions. Specifically, the authors consider a compartmental model consisting of four populations: uninfected CD4$^+$ cells, $x(t)$, infected CD4$^+$ cells, $y(t)$, HTLV-I-specific CTLs, $z(t)$, and the level of IFNγ, $r(t)$. The illustration in Figure 2.5 shows a schematic diagram of the interactions that are represented by the model.

Letting $'$ denote the derivative with respect to time, this leads to the following system of differential equations:

$$\begin{aligned} x' &= \lambda - dx - \beta xy \\ y' &= \beta xy + sy - \varepsilon zy \\ z' &= czy - b_2 z \\ r' &= hf(y)z - gr, \end{aligned} \tag{2.2}$$

where $f(y)$ is a function determining the magnitude of IFNγ production which, in numerical simulations, has been represented using a Hill function, although the authors remarked that the exact functional form of $f(y)$ does not affect the qualitative results of the model (Asquith and Bangham, 2000).

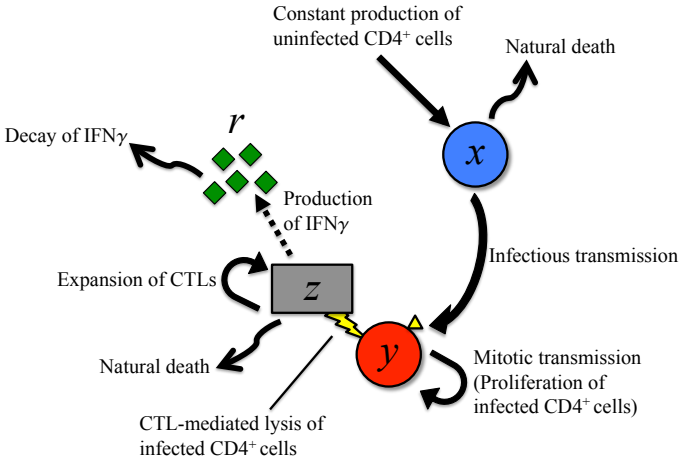


Figure 2.5: A schematic diagram of the model by Asquith and Bangham (2000) that considers the interactions among healthy target cells, x, infected target cells, y, and virus-specific CTLs, z, as well as incorporating the production of inflammatory cytokines IFNγ, denoted by r. The role of target cell latency is not considered.

The authors set out to elucidate the role of CTL-mediated immunity in HTLV-I infection and the impact of CTLs on HTLV-I-associated pathogenesis by incorporating and interpreting the effects of cytokine production from experimental observations. Significantly, the results from the model help unify the two seemingly contradictory viewpoints that CTLs can both prevent and contribute to the progression of disease, explain the observed correlation between proviral load and pathology, and propose new predictions that are experimentally testable.

The work by Asquith and Bangham (2000) illustrates how mathematical modelling can be used to provide plausible explanations for existing experimental or clinical data as well as stimulate the development of new biological hypotheses, and reminds us of the benefits of linking theoretical and empirical tools to advance our understanding of complex systems.

2.7.3 Gómez-Acevedo, Li, and Jacobson (2010)

Gómez-Acevedo et al. (2010) studied a mathematical model based on that of Wodarz et al. (1999) which focuses on providing a complete mathematical analysis of the dynamics of the HTLV-I-specific CTL response. The authors consider three populations of cells: Healthy CD4$^+$ helper T-cells $x(t)$, infected CD4$^+$ helper T-cells $y(t)$, and HTLV-I-specific CD8$^+$ CTLs $z(t)$. The production of new target cells occurs at a constant rate λ. Infectious transmission via direct cell-to-cell contact is given by the term $\beta x(t)y(t)$ and infected cells are subject to strong adaptive immune responses. The effects of mitotic transmission are ignored. The humoral, or antibody response, is assumed to rapidly eliminate newly infected cells so that only a fraction $\sigma\beta x(t)y(t)$, where $0 \leq \sigma \leq 1$, survive. Cellular immunity,

driven by anti-HTLV-I CTLs, targets and lyses infected cells at a rate $\gamma y(t)z(t)$. A key feature of the model is the consideration of a more general function for CTL proliferation motivated by the biological justification that by reducing the proviral load, anti-HTLV-I CTLs simultaneously reduce the amount of antigen that stimulates their own proliferation. Hence, the authors assume that CTL expansion takes the density-dependent form $\nu y(t)z(t)/\big(z(t)+K\big)$, where $K > 0$. Natural death rates of healthy and infected target cells are $\mu_1 x(t)$ and $\mu_2 y(t)$, respectively, and the decay of CTLs is $\mu_3 z(t)$. Figure 2.6 shows a diagrammatic representation of the model.

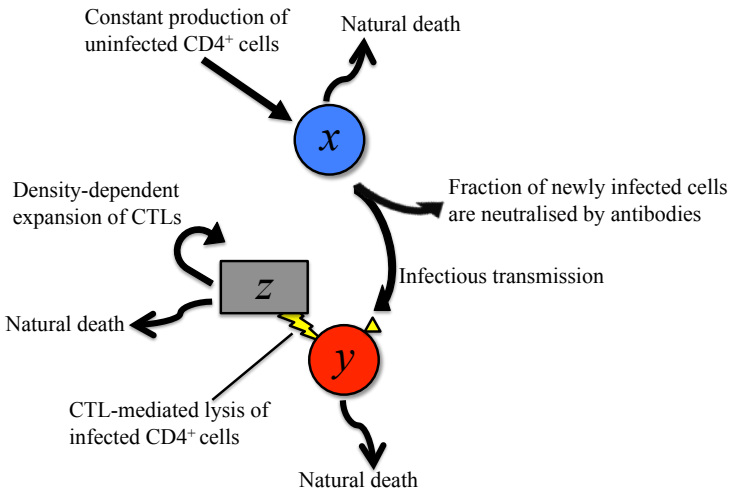


Figure 2.6: A schematic diagram of the model by Gómez-Acevedo et al. (2010) that considers the interactions among healthy target cells, x, infected target cells, y, and virus-specific CTLs, z. Mitotic transmission of the provirus and target cell latency are not considered.

Denoting by $'$ the derivative with respect to time, the resulting system of differential equations is:

$$\begin{aligned} x' &= \lambda - \mu_1 x - \beta xy \\ y' &= \sigma\beta xy - \mu_2 y - \gamma yz \\ z' &= \nu\frac{yz}{z+K} - \mu_3 z. \end{aligned} \tag{2.3}$$

With non-negative initial conditions, Gómez-Acevedo et al. (2010) show that system (2.3) has two different types of endemic equilibria, one corresponding to chronic infection without a CTL response and the other corresponding to chronic infection that stimulates and maintains a persistent CTL response. The authors interpret the former state as that of an asymptomatic carrier and the latter state as that of a HAM/TSP patient. Each of the two chronic infection steady states are shown to be globally stable within respective parameter regions dependent upon two threshold parameters (which the authors refer to as R_0 and

R_1, where $R_1 < R_0$). An important finding from this model is the identification of a control strategy to prevent disease by limiting the CTL response, whose chronic activation has been postulated to cause the damage seen in HAM/TSP patients (Biddison et al., 1997; Jacobson, 2002; Mosley and Bangham, 2009).

The model by Gómez-Acevedo et al. (2010) reinforces the need to consider the CTL response in the dynamics of HTLV-I infection. Nevertheless, further investigation is required to shed light on the processes that HTLV-I uses to evade host immunity while harbouring a high proviral load.

2.7.4 Li and Lim (2011)

Li and Lim (2011) considered the dynamic interactions that occur in the pool of $CD4^+$ helper T-cells, the primary targets of HTLV-I infection, consisting of both uninfected (i.e. healthy) and infected cells. Healthy cells are denoted by $x(t)$. One key aspect of this model is the separation of infected target cells into two distinct compartments depending on the state of activation, which is assumed to be associated with the expression of viral proteins: latently infected cells, $u(t)$, are resting proviral cells that hide viral expression, and actively infected cells, $y(t)$, are activated proviral cells that persistently express viral antigens. An illustration of the interactions captured by the model is shown in Figure 2.7.

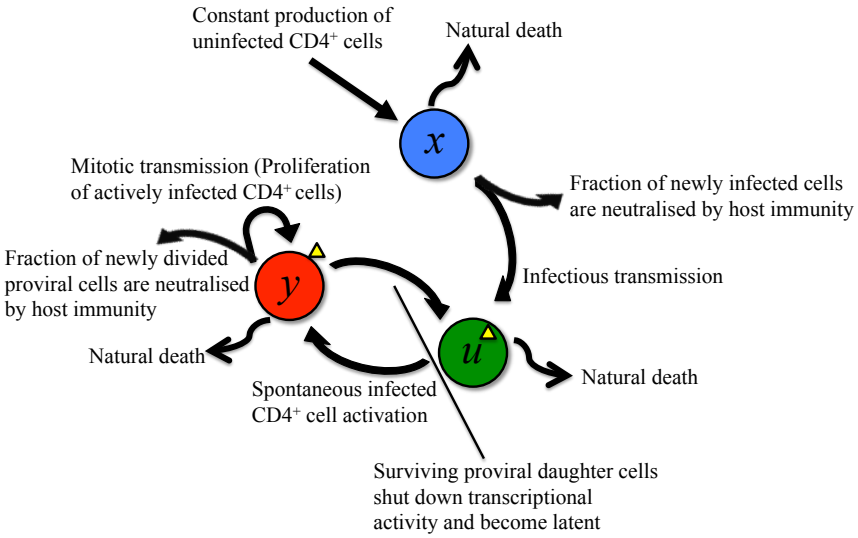


Figure 2.7: A schematic diagram of the model by Li and Lim (2011) that considers the interactions among healthy target cells, x, latently infected target cells, u, and actively infected target cells, y. An explicit incorporation of the virus-specific CTL response is not considered.

Letting $'$ denote the derivative with respect to time, the resulting system of equations

is:

$$\begin{aligned} x' &= \lambda - \beta xy - \mu_1 x \\ u' &= \sigma\beta xy + \epsilon ry\left(1 - \frac{x+u}{k}\right) - (\tau + \mu_2)u \\ y' &= \tau u - \mu_3 y, \end{aligned} \tag{2.4}$$

which is considered with non-negative initial conditions.

Because this model is important for the work in the following chapters, we briefly explain the meaning of each term in the equations. Healthy (uninfected) $CD4^+$ T-cells are produced at a constant rate λ. Infectious spread of the provirus occurs by direct cell-to-cell contact between an actively infected target cell and a healthy target cell at a rate β and is assumed to follow mass-action kinetics. Actively infected target cells also undergo mitotic division at a rate r. Due to the biologically motivated assumption that the vast majority of infected target cells are latent at any given time, the authors describe the mitotic spread of the provirus using a modified logistic growth term $ry\big(1-(x+u)/k\big)$, where k is the carrying capacity of the $CD4^+$ helper T-cell population. Although the immune system dynamics are not modelled explicitly, the authors incorporate a static effect of host immunity. Specifically, the parameters σ and ϵ, where $\sigma, \epsilon \in (0, 1)$, represent proportions of newly infected cells from infectious and mitotic transmission, respectively, that survive immune-mediated elimination. These cells are then assumed to silence viral expression and enter the latently infected target cell compartment. Meanwhile, latently infected cells return to an activated state at a rate τ. Lastly, each of the target cell populations (i.e. healthy, latently infected, and actively infected) are removed from the system via natural death according to exponential decay at respective rates μ_1, μ_2 and μ_3. All parameter values are assumed to be positive.

The authors demonstrate that system (2.4) exhibits the phenomenon of a backward bifurcation resulting in a region of bi-stability with respect to the basic reproduction number for viral infection R_0. Specifically, there is a range of the parameter R_0 for which multiple stable equilibria co-exist. Backward bifurcation and bi-stability suggest that chronic HTLV-I infection is difficult to clear due to hysteresis effects.

The consideration of a latent compartment of proviral cells in model (2.4) allows for a closer examination of the issue of viral persistence. The authors focus their investigations on the role of the immunodominant viral protein Tax, whose spontaneous expression is assumed to coincide with infected target cell activation. They demonstrate that in the chronic phase of infection, although the majority of the proviral load is comprised of latent proviral cells, a chronic infection equilibrium state includes Tax-expressing proviral cells and must be reached through a dynamic balance between infected target cell latency and activation. One main implication of this approach is the identification of the rate of Tax

expression, given by the parameter τ in model (2.4), as having a substantial impact on the outcome of HTLV-I infection. Namely, spontaneous viral expression is a crucial factor that drives chronic infection, determines the proviral load, and affects the time it takes for an individual to reach the equilibrium state. These observations suggest that this aspect may be useful to consider in future mathematical models of HTLV-I infection.

As pointed out earlier, the model by Li and Lim (2011) incorporates the effects of adaptive immunity on proviral cells implicitly by considering surviving fractions of infected cells from infectious and mitotic transmission routes. An explicit compartment of HTLV-I-specific CTLs is not taken into account. In the following chapter, we begin the first stage of our work by extending this model and constructing a new mathematical model to describe the infection of $CD4^+$ helper T-cells which takes into account a dynamically changing CTL response. This new model will be referred to as the *baseline model* and it forms the foundation of the work presented in this thesis.

Chapter 3

Constructing a Baseline Mathematical Model

In this chapter we construct and analyse a baseline mathematical model that captures the essential elements of HTLV-I infection *in vivo* and lays down the foundation for subsequent chapters[1]. In Section 3.1, we begin by posing three key questions in current HTLV-I research which motivate the formulation of our mathematical model. Then, in Section 3.2, we present our mathematical model, which is given by the four-dimensional non-linear system of ordinary differential equations (3.1) describing a dynamically changing within-host environment involving the interactions among healthy, latently infected, and actively infected target cells, and anti-HTLV-I CTLs. Next, in Section 3.3, we consider each parameter in the mathematical model in turn and discuss biologically reasonable ranges in which they lie from experimental and theoretical studies on T-cell dynamics and the kinetics of HTLV-I infection. These parameter values are adopted in numerical simulations in this chapter as well as in subsequent chapters. Following this, in Section 3.4, we discuss and establish the global qualitative behaviour present in the model, which relies on dynamical systems theory and Lyapunov functions to prove stability results. In our mathematical analysis, we first determine a feasible region which bounds the solutions of our model, whose corresponding biological interpretation is that T-cell populations can neither be negative, nor can they become unbounded. We then proceed to characterise the global asymptotic behaviour of solutions, and identify the basic reproduction number for viral infection, R_0, as a sharp threshold parameter that completely determines the global dynamics of the system: If $R_0 < 1$, then the virus is cleared. If $R_0 > 1$, then the infection becomes chronic. A complete understanding of the global behaviour of our model is important as it allows us to study the finer aspects of our model in a robust manner as we use the model to gain some biological insights. In Section 3.5 we non-dimensionalise and re-scale model (3.1)

[1]The work presented in this chapter has been published (Lim and Maini, 2014)

under biologically realistic parameter values to characterise the transient behaviour of the system using asymptotic analysis. The remainder of the chapter focusses on the biological applications and outcomes of the baseline model. In Section 3.6 we distinguish between two important measures of viral detriment which our modelling approach allows us to consider, namely the proviral load and the activated proportion of infected cells. Following this, we proceed to address the three key questions of interest here. Namely, in Section 3.7, we explore the effects of spontaneous viral expression that accompanies infected T-cell activation to understand why HTLV-I is not a completely silent infection. Then in Section 3.8, we consider factors that define an effective virus-specific cellular immune response and speculate why infection ultimately persists despite the presence of anti-viral host immunity. In Section 3.9, we examine our mathematical model in the context of pathology. Results from our investigations provide insights regarding HTLV-I-associated disease and suggest a novel hypothesis for pathogenesis. Penultimately, in Section 3.10, we summarise our principal mathematical results as well as provide biologically meaningful answers arising from exploration of model (3.1) in response to the motivating questions posed. Lastly, in Section 3.12, we conclude the chapter with several final remarks on the baseline model.

3.1 Motivation

The formulation of a baseline mathematical model is principally motivated by the desire to elucidate the following questions.

(1) Since viral activation exposes infected cells to immune-mediated surveillance, why is HTLV-I not silent? In other words, what benefit does the HTLV-I provirus gain in becoming activated and expressing viral antigens?

(2) What determines the strength of the HTLV-I-specific CTL response, and why is infection life-long? That is, how can we evaluate the quality of the HTLV-I-specific CTL response, and why does host immunity fail to eradicate the virus?

(3) What characterises HTLV-I-linked pathology? Indeed, how can one determine the severity of viral detriment considering that the size of the proviral load is insufficient to classify clinical status? Can a mathematical modelling approach suggest an alternative, quantifiable measure to determine the outcome of the infection? If such a measure does exists, what implications might it have on the diagnosis, development, and treatment of HTLV-I-associated disease?

3.2 Baseline Mathematical Model Formulation

Relatively recently, Asquith and Bangham (2008) have proposed an experimental hypothesis for the infection and persistence of HTLV-I *in vivo* which motivated the formulation of a mathematical model by Li and Lim (2011) illustrating the balance between latency and activation in the target cell dynamics of the viral infection. In this model, the action of anti-viral host immunity was considered implicitly via fractions of newly infected cells that survive elimination. However, the pool of HTLV-I-specific immune effectors is highly dynamic and continuously changing in response to the virus population.

As cellular immunity is widely believed to be the most significant factor in determining the outcome of infection, we extend the model of Li and Lim (2011) to incorporate the two key features of HTLV-I infection *in vivo*: (i) viral latency, and (ii) the HTLV-I-specific CTL response. The separation of infected target cells into two distinct compartments, latently infected and actively infected, is crucial to understanding the persistence of HTLV-I in the midst of a chronically stimulated anti-viral host immune response.

The primary target cells of HTLV-I infection are $CD4^+$ helper T-cells, which we will initially separate into three different compartments. To model the explicit role of anti-viral cellular immunity, we also include an additional compartment of HTLV-I-specific $CD8^+$ CTLs. In particular, we define

$x(t)$: density of healthy $CD4^+$ helper T-cells at time t,

$u(t)$: density of latently infected $CD4^+$ helper T-cells at time t,

$y(t)$: density of actively infected $CD4^+$ helper T-cells at time t,

$z(t)$: density of HTLV-I-specific $CD8^+$ CTLs at time t.

We consider a mathematical model of HTLV-I infection given by the following four-dimensional non-linear system of ordinary differential equations:

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \overbrace{\lambda}^{\text{T-cell production}} - \overbrace{\beta xy}^{\text{infectious transmission}} - \overbrace{\mu_1 x}^{\text{natural death}} \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= \overbrace{\beta xy}^{\text{infectious transmission}} + \overbrace{ry}^{\text{mitotic transmission}} - \overbrace{(\tau+\mu_2)u}^{\text{spontaneous activation and natural death}} \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \overbrace{\tau u}^{\text{spontaneous activation}} - \overbrace{\gamma yz}^{\text{CTL-mediated lysis}} - \overbrace{\mu_3 y}^{\text{natural death}} \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= \overbrace{\nu y}^{\text{CTL proliferation}} - \overbrace{\mu_4 z.}^{\text{natural death}}
\end{aligned} \tag{3.1}$$

A schematic representation of the biological mechanism of HTLV-I infection *in vivo* upon which our model is based is shown for reference in Figure 3.1.

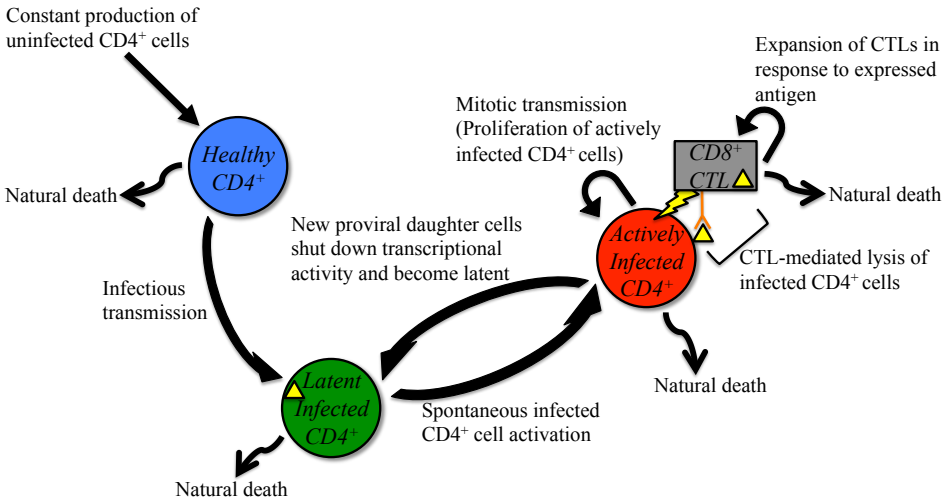


Figure 3.1: A schematic representation of the biological mechanism of HTLV-I infection *in vivo* that motivates the formulation of model (3.1). Healthy $x(t)$, latently infected $u(t)$, and actively infected $y(t)$ target cells are represented by blue, green, and red circles, respectively, while HTLV-I-specific CTLs $z(t)$ are represented by the grey rectangle.

In fact, model (3.1) can be characterised by a set of seven distinct types of reactions that take place among the various T-cell populations. These reactions are motivated by within-host immunological processes which we will justify one by one in the following paragraph. Briefly, the reaction kinetics can be illustrated as below. Short comments are written in square brackets.

(i) Production of healthy target cells from thymus:

$$\emptyset \xrightarrow{\lambda} x, \qquad \text{[Susceptible cells are produced at a constant rate.]}$$

(ii) Infectious or horizontal transmission of provirus:

$$x + y \xrightarrow{\beta} u + y, \qquad \text{[Newly infected cells silence viral expression.]}$$

(iii) Mitotic or vertical transmission of provirus:

$$y \xrightarrow{r} u + y, \qquad \text{[Division is asymmetric.]}$$

(iv) Spontaneous infected T-cell activation:

$$u \xrightarrow{\tau} y, \qquad \text{[Latently infected cells become activated and express viral proteins.]}$$

(v) CTL-mediated lysis of activated infected cells:

$$y + z \xrightarrow{\gamma} z, \qquad \text{[Host immunity only recognises activated infected cells.]}$$

(vi) Proliferation of CTLs in direct response to activated infected cells:

$$y \xrightarrow{\nu} y+z, \qquad \text{[The presence of expressed viral antigen stimulates CTL growth.]}$$

(vii) Natural death of all cell populations:

$$x \xrightarrow{\mu_1} \emptyset, \quad u \xrightarrow{\mu_2} \emptyset, \quad y \xrightarrow{\mu_3} \emptyset, \quad z \xrightarrow{\mu_4} \emptyset. \qquad \text{[The rate of decay is constant.]}$$

Let us now motivate each term individually. Naïve $CD4^+$ helper T-cells are produced in the bone marrow, then migrate to the thymus where they mature before being released into the peripheral blood (Parham, 2005). These mature $CD4^+$ helper T-cells are initially uninfected and we describe their rate of entry into the periphery by a constant, λ [Reaction (i)]. There are two routes of transmission for HTLV-I within the target cell population: horizontal or 'infectious', and vertical or 'mitotic'. Infectious transmission requires direct contact between an actively infected and a healthy target cell, and involves a restructuring of the cell cytoskeleton to create a tight junction known as a virological synapse, across which the viral genome is transported from the infected target cell to the uninfected one (Igakura et al., 2003; Shiraki et al., 2003). Upon infection, the newly infected target cell silences viral expression and becomes latently infected. The precise mechanisms behind the suppression of viral proteins are not yet understood (Asquith and Bangham, 2008). We describe the cell-to-cell virus transmission using a bilinear incidence term $\beta x(t)y(t)$ [Reaction (ii)], where β is the coefficient of infectious transmissibility. The mitotic route of transmission occurs when an actively infected target cell divides, creating two genetically identical daughter cells (Bangham, 2000a). One way to represent the mitotic route of viral transmission is to assume that the proliferation of actively infected cells follows a logistic growth pattern given by a term $ry(t)\big(1 - \big(x(t) + u(t) + y(t)\big)/k\big)$, where r is the rate of rapid selective division, and k is the carrying capacity of $CD4^+$ helper T-cells. However, from studies of T-cell dynamics, the proliferation and removal rates of $CD4^+$ helper T-cells have been quantified (Asquith et al., 2007; Kirschner and Webb, 1996; Nelson et al., 2000), and suggest that the maximum proliferation rate is in general less than the removal rates due to natural death. Thus, it is plausible that even in the presence of rapid selective mitotic division, target cell populations do not exceed the total $CD4^+$ helper T-cell carrying capacity. We have investigated numerically the dynamics of the model using a full logistic term and observed no qualitative difference in the behaviour of trajectories (see Figures C.1 and C.2 in Appendix C.1 for details). Hence, with respect to our model we will assume that $x(t)+u(t)+y(t) \ll k$ for all $t \geq 0$, so that the proliferation of actively infected target cells follows an exponential growth profile. As a result, we choose to represent infected T-cell proliferation

using a linear growth term $ry(t)$ rather than a logistic growth term [Reaction (iii)]. As with infectious transmission, newly infected target cells arising from mitotic division immediately hide the expression of viral genes and enter the latently infected cell compartment, thereby evading the immune system. At this point, we remark that although mitosis is a process that occurs in all target cells, it has been experimentally shown that the rapid rate of infected target cell proliferation is significantly faster than normal homeostatic division (Asquith et al., 2007), but only so during intermittent expression of the provirus. The set of such provirus-expressing infected target cells coincides in our model with the pool of actively infected target cells. Latently infected target cells do not express the provirus and are assumed to proliferate at the same rate as normal, healthy $CD4^+$ helper T-cells. To avoid unnecessarily complicating the model equations, we do not consider the passive proliferation of healthy and latently infected target cells.

Latency allows the proviral cell to escape lysis by anti-HTLV-I CTLs, but hinders both the infectious and mitotic routes of viral transmission. Meanwhile, activation promotes viral transmission, but simultaneously exposes the proviral cell to immune surveillance (Asquith and Bangham, 2008). It is becoming increasingly clear that a dynamic balance between infected target cell latency and activation exists. Indeed, while the vast majority of proviral cells are latent at any given time, it has been observed experimentally that a small proportion of latently infected $CD4^+$ helper T-cells spontaneously express viral proteins and become actively infected (Asquith et al., 2000; Asquith and Bangham, 2008; Asquith et al., 2007). We represent the transition of a proviral cell from the latently infected state to the actively infected state by a term $\tau u(t)$ [Reaction (iv)], where τ is the rate of spontaneous infected T-cell activation or viral protein expression. In the activated state, proviral cells are subject to strong selection by HTLV-I-specific CTLs that recognise expressed HTLV-I antigens, such as epitopes on the immunodominant virus protein Tax and the minus-strand viral gene product HBZ (Gaudray et al., 2002; Goon et al., 2004a). We model the elimination of actively infected target cells by virus-specific CTLs using a bilinear incidence term $\gamma y(t)z(t)$ [Reaction (v)], where γ represents the rate of CTL-mediated lysis. The pool of CTLs is maintained by antigenic stimulation from virus-expressing activated proviral cells (Asquith et al., 2007; Bangham et al., 2009). The replenishment of these CTLs is described by the term $\nu y(t)$ [Reaction (vi)], where ν is the rate of CTL proliferation or turnover, also referred to as the CTL responsiveness (Nowak and Bangham, 1996; Wodarz et al., 2001).

Lastly, all T-cell populations under consideration, including $CD4^+$ helper T-cells and $CD8^+$ cytotoxic T-lymphocytes (CTLs), are removed from the system via natural cell death. We represent the removal of each cell type by a rate proportional to its density. Healthy,

latently infected, and actively infected target cells die at respective rates μ_1, μ_2, and μ_3, and the removal rate of $CD8^+$ CTLs is μ_4 [Reaction (vii)].

All parameters are assumed to be positive. To close the model, we need to specify the initial conditions, and this is done in the next section.

Remark 3.1. It should be clarified that at a more detailed level, a distinction exists between two subclasses of $CD4^+$ helper T-cells in the host-virus interaction above. In particular, the production of T-cells arising from thymic exit describes the subclass of naïve $CD4^+$ T-cells. Meanwhile, it has been demonstrated experimentally that the pool of memory $CD4^+$ T-cells, namely those that are specific to HTLV-I, are the ones most susceptible to infection by the virus (Goon et al., 2004b). An implication of this observation is that our baseline mathematical model (3.1) is likely over-estimating the number of infected cells that occur by infectious cell-to-cell transmission. However, in reality it is not known what proportion of the $CD4^+$ T-cell population are specific to HTLV-I, thus raising a challenge in differentiating between the two subclasses of $CD4^+$ helper T-cells. Because we are primarily interested in examining the overall dynamics of persistent viral infection, we avoid over-complicating our mathematical model at this stage and, as a simplification, we designate the 'target cell' population to comprise memory $CD4^+$ helper T-cells, with the naïve assumption that such T-cells exit the thymus and are also infected by the virus.

3.3 Parameter Values and Numerical Simulations

In this section, we discuss the parameter values that we use to simulate the baseline model (3.1), which have been estimated using both experimental and theoretical methods in studies of $CD4^+$ lymphocyte kinetics by Kirschner and Webb (1996), and Nelson et al. (2000). Asquith et al. (2005b, 2007) have also quantified the *in vivo* kinetics of $CD4^+$ helper T-cells and $CD8^+$ CTLs in the context of persistent infection by HTLV-I in both asymptomatic carriers and in HAM/TSP patients. Using these biologically realistic parameters, we then simulate the time evolution graphs for model (3.1) numerically, which allows us to identify some of the possible dynamical behaviours present and get a sense of the kinds of mathematical techniques we might be able to apply to analyse the model.

3.3.1 Parameter Values

The production, λ, of healthy, naïve $CD4^+$ helper T-cells exiting the thymus falls in the range of 0–10 cells/mm^3/day (Kirschner and Webb, 1996; Perelson et al., 1993), and we

select a value of 10 cells/mm^3/day*. As infection by HTLV-I only causes minor detriment to T-cell functionality (Asquith and Bangham, 2007), it is expected that all three populations of target cells considered in our model display natural death rates similar to those of healthy target cells, between 0.01–0.11 day^{-1} (Kirschner and Webb, 1996; Nelson et al., 2000; Ribeiro et al., 2002). Extracting from studies of HIV-1 infection (Ribeiro et al., 2002), it is assumed that activated CD8$^+$ CTLs die at similar rates in both uninfected and infected individuals, with the death rate lying between 0.03–0.05 day^{-1}. Experiments have measured the rate of rapid Tax-driven selective mitosis r to lie within the range 0.01–0.045 day^{-1} (Asquith et al., 2007; Kirschner and Webb, 1996), with higher rates of CD4$^+$ helper T-cell turnover usually associated with disease status, e.g. HAM/TSP. Because our model does not distinguish between the subclasses of naïve and memory CD4$^+$ T-cells, for the model simulations, we have chosen a proliferation rate of 0.011 day^{-1} to coincide with the maximal proliferation rate of CD4$^+$ cells as measured by Kirschner and Webb (1996).

We consider values for the coefficient of infectious transmissibility β to be on the order of 10^{-3} mm^3/cell/day, which is consistent with those used in recent mathematical models of HTLV-I infection (Gómez-Acevedo and Li, 2005; Gómez-Acevedo et al., 2010; Li and Lim, 2011). Asquith et al. (2007) have quantified the rate of expression, τ, of Tax in proviral cells to be between 0.0003–0.03 per day, and an intermediary value of 0.003 per day is chosen. The rate of lysis by anti-HTLV-I CTLs γ in asymptomatic carriers of the virus depends on each individual, and has been measured to be between 0.007–0.220 per CD8$^+$ cell per day with a median value of 0.029 day^{-1} (Asquith et al., 2005b). Finally, the proliferation rate of HTLV-I-specific CD8$^+$ CTLs has been measured to be in the range 0.009–0.161 per day with a median value of 0.036 day^{-1} (Asquith et al., 2007). The biological meaning of the parameters, the primary sources for the parameter ranges, and the specific choices for the dimensional parameter values within the relevant ranges are summarised in Table 3.1.

*Using both experimental techniques and mathematical modelling, Bains et al. (2009) have quantified the rate of thymic export of CD4$^+$ helper T-cells to be 3.5×10^8 CD4$^+$ T-cells per day, with a range between $(0.1, 27) \times 10^8$ CD4$^+$ T-cells by the age of 20. Considering that the blood volume of an average human weighing 70 kg is roughly 5.5 litres, taking our selected value of λ yields 10 cells/mm^3/day $\times$ 5.5×10^6 mm^3 $= 5.5 \times 10^7$ CD4$^+$ T-cells per day, which is within the range of thymic exit estimated by Bains et al. (2009).

Initial Condition	Value		Source
$x(0)$	~ 850 cells/mm^3		(Bofill et al., 1992)
$u(0)$	0.1 cells/mm^3		Assumption on initial infection
$y(0)$	0.5 cells/mm^3		Assumption on initial infection
$z(0)$	0.1 cells/mm^3		Assumption on initial HTLV-I-specific CTL abundance
Dimensional Parameter	**Value**	**Biological Meaning**	**Source**
λ	10 cells/mm^3/day	rate of production of target cells (CD4$^+$ helper T-cells)	(Bains et al., 2009; Kirschner and Webb, 1996; Perelson et al., 1993)
β	0.001 mm^3/cell/day	infectious transmissibility coefficient	(Gómez-Acevedo and Li, 2005; Gómez-Acevedo et al., 2010; Li and Lim, 2011; Perelson, 1989)
r	0.011 day^{-1}	selective proliferation rate of actively infected cells	(Asquith et al., 2007; Kirschner and Webb, 1996)
k	1150 cells/mm^3	target cell carrying capacity	(Gómez-Acevedo and Li, 2005; Li and Lim, 2011)
τ	0.003 day^{-1}	rate of spontaneous Tax expression	(Asquith et al., 2007; Li and Lim, 2011)
γ	0.029 mm^3/cell/day	rate of CTL-mediated lysis of actively infected cells	(Asquith et al., 2005b, 2007)
ν	0.036 day^{-1}	proliferation rate of CTLs (or CTL responsiveness)	(Asquith et al., 2007)
μ_1	0.012 day^{-1}	natural death rate of healthy cells	(Kirschner and Webb, 1996; Nelson et al., 2000; Ribeiro et al., 2002)
μ_2	0.03 day^{-1}	natural death rate of latently infected cells	(Kirschner and Webb, 1996; Nelson et al., 2000; Ribeiro et al., 2002)
μ_3	0.03 day^{-1}	natural death rate of actively infected cells	(Kirschner and Webb, 1996; Nelson et al., 2000; Ribeiro et al., 2002)
μ_4	0.03 day^{-1}	natural death rate of virus-specific CTLs	(Ribeiro et al., 2002)

Table 3.1: Table of biologically relevant dimensional initial conditions and parameter values that describe chronic infection by HTLV-I. In general, median values have been selected. The ranges for each parameter can be found in the text. Values within these ranges are used to parameterise the various HTLV-I models explored in this thesis.

Before we can solve model (3.1), we discuss a biologically reasonable initial state of the system at the onset of infection. In the absence of infection, the normal CD4$^+$ helper T-cell count averages 850 cells/mm^3 of peripheral blood, although it is known that there is a wide variance between individuals, with healthy values in adolescents and adults ranging from

500 cells/mm^3 to 1200 cells/mm^3[†] (Bofill et al., 1992).

We assume that the individual is initially healthy, and that the start of infection occurs after the introduction of a small dose of infected cells, say on the order of 10^{-1} cells/mm^3. Next, it is known that CTLs of a diverse range of specificities are continuously being produced. The average level of CD8$^+$ cytotoxic T-lymphocytes in a normal, healthy individual is roughly 550 cells/mm^3 of peripheral blood (Bofill et al., 1992). We will assume that at the start of infection, the number of circulating CD8$^+$ CTLs in an individual that can recognise HTLV-I epitopes is on the order of 10^{-1} cells/mm^3. Here we remark that although the magnitude of the initial level of viraemia is small relative to the size of the target cell pool, these values, measured in units of cells per mm^3 of peripheral blood, are assumed to be homogeneous throughout the entire body of an individual, whose blood volume is several orders of magnitude higher ($\sim$ 5.5 L of blood for an average human weighing roughly 70 kg). Thus, stochastic effects are negligible and we are justified in adopting a deterministic modelling approach.

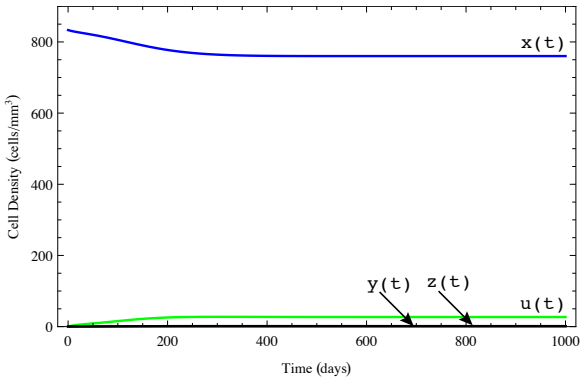
3.3.2 Numerical Simulations

To gain some insight as to how we will approach the mathematical analysis of model (3.1), we use these biologically motivated initial conditions and parameter values as in Table 3.1, and run preliminary numerical simulations of the model. Using the built-in differential equation solver in Mathematica (version 10.0.0.0), we solve model (3.1) numerically and plot the solution as in Figure 3.2. The `NDSolve` command in Mathematica automatically selects a numerical algorithm to optimally solve the designated system of ordinary differential equations, and includes both explicit and implicit methods such as Euler, Adams, and Runge-Kutta of arbitrary order. Figure 3.2(a) shows all four cell populations and highlights their relative abundances. As the magnitudes of the terms in the u, y, z-equations are much smaller than those of the x-equation, we take a closer look at their behaviours by magnifying the vertical axis, roughly 30 times in Figure 3.2(b) and 600 times in Figure 3.2(c). We observe that after some time, the solution appears to settle at a positive equilibrium which may be reached by damped oscillations. In the following section, we will examine the underlying dynamical behaviour of model (3.1) in more detail.

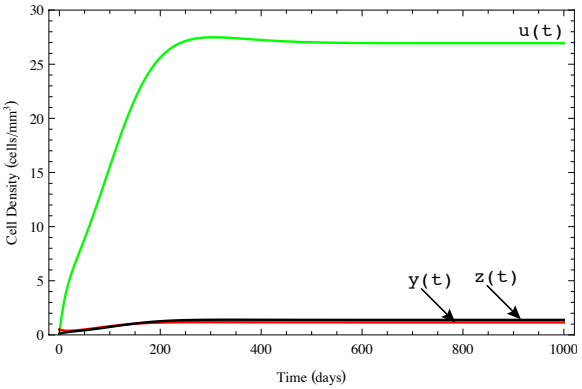
3.4 Global Qualitative Behaviour

The purpose of this section is to provide a complete mathematical description of the global qualitative behaviour of all possible solutions of model (3.1). Establishing the global dy-

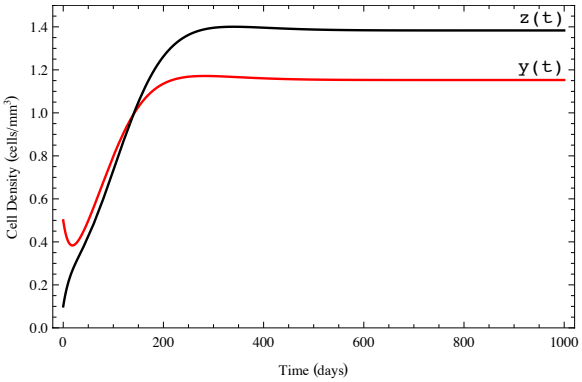
[†]For reference, Acquired Immune Deficiency Syndrome (AIDS) caused by the human immunodeficiency virus (HIV) is characterised by a CD4$^+$ helper T-cell count below 200 cells/mm^3.



(a) Numerical solution of model (3.1).



(b) Zooming in on the behaviour of $u(t)$.



(c) Zooming in on the behaviour of $y(t)$ and $z(t)$.

Figure 3.2: Numerical simulation of the solution of model (3.1) with initial conditions and parameter values as in Table 3.1. The level of healthy ($x(t)$), latently infected ($u(t)$), and actively infected ($y(t)$) target cells are shown in blue, green, and red, respectively, and the level of HTLV-I-specific CTLs ($z(t)$) is shown in black. (a) The full dimensional range showing all cell populations and highlighting their relative abundances. (b)–(c) A closer examination of the behaviour of $u(t)$, $y(t)$ and $z(t)$ in model (3.1).

namics of our mathematical model means that we have a complete understanding of the qualitative behaviour of solutions of the model. This puts us in a position where we can fully explore the model without having to worry about the occurrence of unexpected phenomena, and allows us to draw robust conclusions from further investigations of the model in the remainder of this chapter. More broadly, a well-defined mathematical model can also help us identify specific processes that can lead to experimental designs which test the validity of the model representation we have chosen and determine its applicability to the biological system under study.

3.4.1 Biologically Realistic Region

The first step in the mathematical analysis of model (3.1) is to determine a bounding region for the T-cell populations to ensure that our model is biologically reasonable. In particular,

T-cell populations should never become negative, nor should they become unbounded. Our first result, Theorem 3.1, formalises such a notion.

Denote by $\mathbb{R}^4_+$ the closed positive orthant of $\mathbb{R}^4$. Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$, and consider the closed, bounded region

$$\Gamma := \left\{ (x, u, y, z) \in \mathbb{R}^4_+ : x \leq \frac{\lambda}{\mu_1},\ x + u + y \leq \frac{\lambda}{\mu - r},\ z \leq \frac{\lambda\nu}{\mu_4(\mu - r)} \right\}.$$

Theorem 3.1. *Assuming that $r < \mu$, the set Γ is positively invariant with respect to model (3.1). All solutions are bounded for $t \geq 0$ and eventually enter Γ.*

Proof of Theorem 3.1. It can be seen from the model equations that for any set of non-negative initial conditions, $x(t), u(t), y(t), z(t) \geq 0$ for all $t \geq 0$ (this is because the equations in model (3.1) do not allow any of the T-cell compartments to become negative when any one of the populations sit on the boundary at zero). It is clear that

$$\frac{\mathrm{d}x}{\mathrm{d}t} \leq \lambda - \mu_1 x,$$

and this implies that

$$\limsup_{t\to\infty} x(t) \leq \frac{\lambda}{\mu_1}.$$

Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$ and suppose that $r < \mu$. Adding the first three equations of model (3.1) yields the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}(x + u + y) \leq \lambda - (\mu - r)(x + u + y),$$

which implies that $\limsup_{t\to\infty}(x + u + y)(t) \leq \lambda/(\mu - r)$. The assumption that $r < \mu$ ensures that the quantity on the right-hand side of the inequality is always positive. Let $\big(x(t), u(t), y(t), z(t)\big)$ be a solution of model (3.1) with $x(0) + u(0) + y(0) \leq \lambda/(\mu - r)$. Then,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \nu y - \mu_4 z \leq \nu\frac{\lambda}{\mu - r} - \mu_4 z \quad \Longrightarrow \quad \limsup_{t\to\infty} z(t) \leq \frac{\lambda\nu}{\mu_4(\mu - r)}.$$

Therefore, we consider the closed, bounded region

$$\Gamma := \left\{ (x, u, y, z) \in \mathbb{R}^4_+ : x \leq \frac{\lambda}{\mu_1},\ x + u + y \leq \frac{\lambda}{\mu - r},\ z \leq \frac{\lambda\nu}{\mu_4(\mu - r)} \right\}.$$

As Γ is positively invariant in $\mathbb{R}^4$ (due to the direction of the vector field for model (3.1) pointing towards and remaining in Γ for every initial element in Γ), model (3.1) is well-posed. This completes the proof of the result. □

In Theorem 3.1 we have assumed that $r < \mu$. This assumption is supported by experimental evidence indicating that the proliferation rate of $CD4^+$ helper T-cells is generally lower than the rate of removal due to natural death (Asquith et al., 2007; Kirschner and Webb, 1996; Nelson et al., 2000). Hence, our results rest on biologically meaningful assumptions and our theoretical result, Theorem 3.1, defines the set Γ as a biologically feasible region on which the model dynamics may be analysed.

3.4.2 Basic Reproduction Number for Viral Infection

The global behaviour of model (3.1) depends crucially on a key parameter, the *basic reproduction number for viral infection*, defined as

$$R_0 = \frac{\tau}{\mu_3(\tau + \mu_2)}\Big(\beta x_H + r\Big), \quad \text{where } x_H = \frac{\lambda}{\mu_1}. \tag{3.2}$$

Biologically, R_0 represents the average number of secondary infected cells produced from a single actively infected cell over its lifetime. There are various ways of deriving the basic reproduction number each one of which, although all yielding the same expression, provides a slightly different perspective to the interpretation of R_0. Three different methods of deriving R_0 are explained in detail in Appendix A.1. We remark that the expression for R_0 derived from model (3.1) displays a similar form to the one determined in the model by Li and Lim (2011).

3.4.3 Characterisation of Global Dynamics

The following theorem summarises our main mathematical result and establishes R_0 as a sharp threshold that characterises the global dynamics of model (3.1) in the biologically realistic region. We present a rigorous mathematical proof of the result, which has been separated into three parts to facilitate ease of understanding. Section 3.4.4 deals with the existence and uniqueness of equilibria, while Sections 3.4.5 and 3.4.6, respectively, demonstrate the stability properties of the infection-free and chronic infection steady states, whenever they exist. We establish the global dynamical behaviour of model (3.1) using the direct method of Lyapunov, a well-known method which has been successfully employed in proving the global stability of equilibria in a variety of population models from epidemiology (Kalajdzievska and Li, 2011; Korobeinikov and Wake, 2002) and immunology (Gómez-Acevedo et al., 2010; Korobeinikov, 2004).

Theorem 3.2 (Global Dynamics of the Baseline Model)**.**

1) *The infection-free equilibrium P_0 always exists. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma}$[‡] and it is globally asymptotically stable in $\bar{\Gamma}$;*

2) *if $R_0 > 1$, then the infection-free equilibrium P_0 is unstable. In addition, a unique chronic infection equilibrium P^* exists in $\mathring{\Gamma}$[§] and it is globally asymptotically stable in $\mathring{\Gamma}$.*

3.4.4 Proof of Theorem 3.2. Part 1: Existence of Equilibria

The first proposition shows that the basic reproduction number for viral infection, given by

$$R_0 = \frac{\tau}{\mu_3(\tau + \mu_2)}\big(\beta x_H + r\big), \quad \text{where } x_H = \frac{\lambda}{\mu_1},$$

as in Equation (3.2), determines the existence of equilibria in $\bar{\Gamma}$.

Proposition 3.1 (Existence of Equilibria)**.**

1) The infection-free equilibrium $P_0 = (x_H, 0, 0, 0)$, where $x_H = \lambda/\mu_1$, always exists in $\bar{\Gamma}$. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma}$;

2) If $R_0 > 1$, there exist exactly two equilibria, P_0 on $\partial\Gamma$ and a unique chronic infection equilibrium $P^* = (x^*, u^*, y^*, z^*)$ in $\mathring{\Gamma}$.

Proof. Equilibria occur when

$$0 = \lambda - \beta xy - \mu_1 x, \tag{3.3a}$$

$$0 = \beta xy + ry - (\tau + \mu_2)u, \tag{3.3b}$$

$$0 = \tau u - \gamma yz - \mu_3 y, \tag{3.3c}$$

$$0 = \nu y - \mu_4 z. \tag{3.3d}$$

Observe that since x, u, y and z are restricted to be non-negative in the biologically feasible region Γ, then from Equations (3.3c) and (3.3d), $u = 0 \iff y = 0 \iff z = 0$ at equilibrium. We will show that there are only two possible non-negative equilibria in our model (3.1): P_0 representing the infection-free state, and P^* representing chronic infection.

[‡]The (topological) closure of Γ in $\mathbb{R}^4$: $\bar{\Gamma} = \Gamma \cup \partial\Gamma$, where $\partial\Gamma$ denotes the boundary of Γ. Note that, in fact, $\Gamma = \bar{\Gamma}$.

[§]The (topological) interior of Γ in $\mathbb{R}^4$: $\mathring{\Gamma} = \bar{\Gamma} \setminus \partial\Gamma$, where $\partial\Gamma$ denotes the boundary of Γ.

We first determine the existence of the *infection-free* equilibrium $P_0 = (x_H, 0, 0, 0)$, where $x_H = \lambda/\mu_1$ is the level of target cells in the absence of an infection and coincides precisely with the normal CD4$^+$ helper T-cell count in a healthy individual.

Next, we search for steady state solutions of the form $P^* = (x^*, u^*, y^*, z^*)$, where $x^*, u^*, y^*, z^* > 0$. We will refer to such a steady state, if it exists, as a *chronic infection* equilibrium. Equations (3.3a) to (3.3d) allow us to express x^*, u^* and y^* in terms of z^*. Namely,

$$x^* = \frac{\lambda\nu}{\beta\mu_4 z^* + \nu\mu_1}, \quad u^* = \frac{\mu_4}{\tau\nu}(\gamma z^* + \mu_3)z^*, \quad y^* = \frac{\mu_4}{\nu}z^*, \tag{3.4}$$

where z^* is a root of the following function $F(z)$:

$$F(z) = \frac{\tau}{\mu_3(\tau + \mu_2)}\left(\frac{\beta\lambda\nu}{\beta\mu_4 z + \nu\mu_1} + r\right) - \frac{\gamma}{\mu_3}z - 1. \tag{3.5}$$

It is possible to solve the equation $F(z) = 0$ analytically. However, we have chosen not to show an explicit formula for z^* as the resulting expression is overly complicated and does not provide any meaningful insight to the task at hand, which is to prove the existence of a unique chronic infection steady state for model (3.1). We instead construct the following argument. Clearly, $F(z)$ is a monotonically decreasing function of z and $\lim_{z\to\infty} F(z) < 0$. Moreover, since

$$F(0) = \big[R_0 - 1\big],$$

then the existence of a unique positive root z^* of $F(z) = 0$ is equivalent to the condition that $R_0 > 1$, thus completing the proof. □

3.4.5 Proof of Theorem 3.2. Part 2: Stability of P_0

The *Jacobian matrix* for model (3.1) is

$$J(x, u, y, z) = \begin{bmatrix} -\beta y - \mu_1 & 0 & -\beta x & 0 \\ \beta y & -\tau - \mu_2 & \beta x + r & 0 \\ 0 & \tau & -\gamma z - \mu_3 & -\gamma y \\ 0 & 0 & \nu & -\mu_4 \end{bmatrix}. \tag{3.6}$$

The local stability of P_0 is given by the following.

Proposition 3.2 (Local Stability of P_0)**.**

i) When $R_0 < 1$, the infection-free equilibrium $P_0 = (x_H, 0, 0, 0)$, where $x_H = \lambda/\mu_1$, is always locally asymptotically stable in the feasible region Γ.

ii) When $R_0 > 1$, P_0 is unstable. Specifically, P_0 is a saddle with $\dim W^s_{loc}(P_0) = 3$ and $\dim W^u_{loc}(P_0) = 1$, where $W^s_{loc}(P_0)$, $W^u_{loc}(P_0)$ denote the local stable and unstable manifolds of P_0, respectively.

Proof. Establishing the local stability of the infection-free equilibrium P_0 is equivalent to determining the signs of the real parts of the eigenvalues of $J(P_0)$, the Jacobian matrix at P_0. The characteristic polynomial is

$$\chi_0(\zeta) = (\zeta + \mu_1)(\zeta + \mu_4)\Big[\zeta^2 + (\tau + \mu_2 + \mu_3)\zeta - \mu_3(\tau + \mu_2)\big[R_0 - 1\big]\Big],$$

where R_0 is the basic reproduction number for viral infection defined in Equation (3.2). The roots of $\chi_0(\zeta)$ are precisely the eigenvalues of the Jacobian matrix $J(P_0)$. They are given by $\zeta_i \in \mathbb{C}$ such that $\chi_0(\zeta_i) = 0$:

$$\zeta_1 = -\mu_1, \quad \zeta_2 = -\mu_4,$$

and

$$\zeta_{3,4} = -\frac{1}{2}(\tau + \mu_2 + \mu_3) \pm \frac{1}{2}\sqrt{(\tau + \mu_2 + \mu_3)^2 + 4\mu_3(\tau + \mu_2)[R_0 - 1]}.$$

Clearly, both $\zeta_1, \zeta_2 < 0$. Moreover, if $R_0 < 1$, then $\mathrm{Re}(\zeta_3), \mathrm{Re}(\zeta_4) < 0$. Hence P_0 is either a stable node or a stable spiral. However, if $R_0 > 1$, then it is readily seen that $\mathrm{Re}(\zeta_3) > 0$, so that P_0 is unstable. Specifically, P_0 is a saddle whose local stable manifold $W^s_{loc}(P_0)$ has dimension 3 and local unstable manifold $W^u_{loc}(P_0)$ has dimension 1. If $R_0 = 1$, then $\zeta_3 = 0$ is a zero eigenvalue of $J(P_0)$ and no immediate conclusions about the local stability of P_0 may be inferred. The preceding argument establishes the basic reproduction number for viral infection R_0 as a threshold parameter that characterises the local stability of the infection-free equilibrium P_0. □

In fact, a stronger statement can be made about the stability of P_0 when $R_0 < 1$, namely that it is globally asymptotically stable in the feasible region Γ.

Proposition 3.3 (Global Stability of P_0)**.** *When $R_0 < 1$, the infection-free equilibrium P_0 is globally asymptotically stable in the feasible region Γ.*

Proof. Assume that $R_0 < 1$. Then by Proposition 3.1, the infection-free steady state P_0 is the only equilibrium point in $\bar{\Gamma}$. Note that $x \leq x_H$ for all $x \in \Gamma$.

Consider

$$L = L(x, u, y, z) = \left(x - x_H - x_H \log \frac{x}{x_H}\right) + u + \frac{\tau + \mu_2}{\tau} y + \frac{\gamma(\tau + \mu_2)}{2\tau\nu} z^2.$$

The function $L : \mathbb{R}^4_+ \to \mathbb{R}$ is positive definite[2], as

[2]We refer the reader to Lemma B.1, which shows a proof for the positive definiteness of the function $f(x) = x - \hat{x} - \hat{x}\log(x/\hat{x})$, where $\hat{x}$ is any positive constant.

(i) $L(x_H, 0, 0, 0) = 0$, i.e. $L(P_0) = 0$; and

(ii) $L(x, u, y, z) > 0$, for all $(x, u, y, z) \in \mathbb{R}^4_+, (x, u, y, z) \neq (x_H, 0, 0, 0)$.

The Lyapunov derivative of L is

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t} &= \left(1 - \frac{x_H}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\tau + \mu_2}{\tau}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\gamma(\tau + \mu_2)}{\tau\nu}z\frac{\mathrm{d}z}{\mathrm{d}t} \\ &= \left(1 - \frac{x_H}{x}\right)(\lambda - \mu_1 x) - \beta xy\left(1 - \frac{x_H}{x}\right) + \Big(\beta xy + ry - (\tau + \mu_2)u\Big) \\ &\quad + \frac{\tau + \mu_2}{\tau}(\tau u - \gamma yz - \mu_3 y) + \frac{\gamma(\tau + \mu_2)}{\tau\nu}z(\nu y - \mu_4 z).\end{aligned}$$

Replacing $\lambda = \mu_1 x_H$ and expanding terms yields

$$\begin{aligned}\frac{\mathrm{d}L}{\mathrm{d}t} &= -\frac{\mu_1}{x}(x - x_H)^2 + \frac{\mu_3}{\tau}(\tau + \mu_2)\big[R_0 - 1\big]y - \frac{\gamma\mu_4}{\tau\nu}(\tau + \mu_2)z^2 \\ &\leq 0, \quad \forall (x, u, y, z) \in \Gamma, \quad \text{since } R_0 < 1.\end{aligned}$$

Hence, L is an appropriate Lyapunov function and, moreover, $\mathrm{d}L/\mathrm{d}t(x_H, 0, 0, 0) = L(x_H, 0, 0, 0) = 0$. Since the feasible region Γ is compact and invariant with respect to the vector field of model (3.1), it then follows from LaSalle's invariance principle (LaSalle, 1976) that every trajectory starting in Γ approaches a set M, where M is the largest invariant subset of

$$I = \left\{(x, u, y, z) \in \mathbb{R}^4_+ : \frac{\mathrm{d}L}{\mathrm{d}t}(x, u, y, z) = 0\right\},$$

which, in the present case, is precisely

$$\{x \equiv x_H\} \cap \{y \equiv 0\} \cap \{z \equiv 0\}.$$

To establish the global stability of P_0 when $R_0 < 1$, it remains to show that the set M consists solely of the equilibrium point P_0. Observe that in the set I, the variable u can take on any arbitrary value along the non-negative real line. Let $\omega(t) = \big(x(t), u(t), y(t), z(t)\big)$ be any trajectory of model (3.1) starting in I, so that $x(0) = x_H$, $y(0) = 0$, and $z(0) = 0$. Then, the behaviour of $\omega(t)$ at $t = 0$ is governed by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}u}{\mathrm{d}t} = -(\tau + \mu_2)u, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = \tau u, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = 0.$$

Hence, $x(t) \equiv x_H$, $z(t) \equiv 0$, and

$$u(t) = u(0)e^{-(\tau+\mu_2)t}.$$

The y-compartment varies according to $\mathrm{d}y/\mathrm{d}t = \tau u(0)$ at time $t = 0$. If $u(0) \neq 0$, then $\mathrm{d}y/\mathrm{d}t > 0$, implying that $y(t_0) \neq 0$ for some $t_0 \neq 0$; that is, the trajectory will leave the set I. For the solution $\omega(t)$ to remain in I for all t, we must have $u(0) = 0$ and therefore $u(t) \equiv 0$. Subsequently, solving the initial value problem $\mathrm{d}y/\mathrm{d}t = 0$, $y(0) = 0$ yields $y(t) \equiv 0$. Taken all together, the above argument means that the maximal invariant set in I is comprised only of the single point $P_0 = (x_H, 0, 0, 0)$, i.e. $M = \{P_0\}$. □

3.4.6 Proof of Theorem 3.2. Part 3: Stability of P^*

The final part of the proof of Theorem 3.2 is to demonstrate the global stability of the chronic infection equilibrium P^* whenever it exists in Γ. We therefore establish the following.

Proposition 3.4 (Global Stability of P^*)**.** *When $R_0 > 1$, the chronic-infection equilibrium P^* is globally asymptotically stable in $\mathring{\Gamma}$.*

Proof. Assume that $R_0 > 1$. Then by Proposition 3.1, the chronic infection equilibrium P^* exists in the feasible region Γ. To prove the result, we construct an appropriate Lyapunov function $V : \mathbb{R}^4_+ \to \mathbb{R}$ that will allow us to conclude the global asymptotic stability of P^* in $\mathring{\Gamma}$.

Suppose that $\big(x(t), u(t), y(t), z(t)\big)$ is a solution of model (3.1) and let

$$V = \left(x - x^* - x^* \log \frac{x}{x^*}\right) + \left(u - u^* - u^* \log \frac{u}{u^*}\right) + \frac{\tau + \mu_2}{\tau}\left(y - y^* - y^* \log \frac{y}{y^*}\right) + \frac{\gamma}{\tau\nu}(\tau + \mu_2)z^*\left(z - z^* - z^* \log \frac{z}{z^*}\right).$$

The function $V = V(x, u, y, z)$ is positive definite with respect to the chronic infection equilibrium point $P^* = (x^*, u^*, y^*, z^*)$. The reason for this is because V is a sum of positive definite functions (see Lemma B.1). Taking the Lyapunov derivative along the solution $\big(x(t), u(t), y(t), z(t)\big)$ yields

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}t} &= \left(1 - \frac{x^*}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \left(1 - \frac{u^*}{u}\right)\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\tau + \mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)\frac{\mathrm{d}y}{\mathrm{d}t} \\ &\quad + \frac{\gamma}{\tau\nu}(\tau + \mu_2)z^*\left(1 - \frac{z^*}{z}\right)\frac{\mathrm{d}z}{\mathrm{d}t} \\ &= \left(1 - \frac{x^*}{x}\right)(\lambda - \beta xy - \mu_1 x) + \left(1 - \frac{u^*}{u}\right)\big(\beta xy + ry - (\tau + \mu_2)u\big) \\ &\quad + \frac{\tau + \mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)(\tau u - \gamma yz - \mu_3 y) + \frac{\gamma}{\tau\nu}(\tau + \mu_2)z^*\left(1 - \frac{z^*}{z}\right)(\nu y - \mu_4 z).\end{aligned}$$

By making the appropriate substitutions using the equilibrium equations (3.3a)–(3.3d), we

find that

$$\begin{aligned}\frac{\mathrm{d}V}{\mathrm{d}t} &= \left(1-\frac{x^*}{x}\right)(\mu_1 x^* - \mu_1 x) + \left(1-\frac{x^*}{x}\right)(\beta x^* y^* - \beta x y) \\ &\quad + \left(1-\frac{u^*}{u}\right)\left(\beta x y + r y - \left(\beta x^* y^* + r y^*\right)\frac{u}{u^*}\right) \\ &\quad + \frac{\beta x^* y^* + r y^*}{\tau u^*}\left(1-\frac{y^*}{y}\right)\left(\tau u - \tau u^* \frac{y}{y^*}\right) + \frac{\tau+\mu_2}{\tau}\left(1-\frac{y^*}{y}\right)(\gamma y z^* - \gamma y z) \\ &\quad + \frac{\gamma}{\tau\nu}(\tau+\mu_2) z^*\left(1-\frac{z^*}{z}\right)\left(\nu y - \nu y^* \frac{z}{z^*}\right) \\ &= \mu_1 x^*\left(2 - \frac{x^*}{x} - \frac{x}{x^*}\right) + \beta x^* y^*\left(3 - \frac{x^*}{x} - \frac{x}{x^*}\frac{u^*}{u}\frac{y}{y^*} - \frac{u}{u^*}\frac{y^*}{y}\right) \\ &\quad + r y^*\left(2 - \frac{u^*}{u}\frac{y}{y^*} - \frac{u}{u^*}\frac{y^*}{y}\right) + \frac{\gamma}{\tau}(\tau+\mu_2) y z^*\left(2 - \frac{z}{z^*} - \frac{z^*}{z}\right) \\ &\leq 0.\end{aligned}$$

The final line follows from the inequalities $\sqrt{a_1 a_2} \leq (a_1 + a_2)/2$ and $\sqrt[3]{a_1 a_2 a_3} \leq (a_1 + a_2 + a_3)/3$, where $a_1, a_2, a_3 > 0$; that is, the geometric mean of n positive numbers can be no greater than their arithmetic mean. Hence, the global stability of P^* has been established when $R_0 > 1$. □

3.5 Non-Dimensionalisation and Asymptotic Analysis

In this section, we seek to characterise the transient behaviour of solutions of model (3.1) under biologically realistic parameters as in Table 3.1. To achieve this, we first non-dimensionalise the system, then re-scale the variables and parameters to order 1, and lastly perform an asymptotic expansion to explore the leading order behaviour.

3.5.1 Non-dimensionalisation

For the non-dimensionalisation, we begin by defining new variables

$$X = \frac{x}{x_0}, \quad U = \frac{u}{u_0}, \quad Y = \frac{y}{y_0}, \quad Z = \frac{z}{z_0}, \quad t' = \frac{t}{t_0},$$

where x_0, u_0, y_0, z_0 are the appropriate re-scalings that make all variables of order 1. We remark that the observation from Figures 3.2(a)–(c) with respect to the relative sizes of x, u, y, z suggest that the ratios u_0/x_0, y_0/x_0 and z_0/x_0 are 'small'. We will make this statement more precise in the following subsection about re-scaling variables to order 1 (refer to Subsection 3.5.2). Substituting into the equations in model (3.1) yields the following non-

dimensional system:

$$\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t'} &= \lambda_N - \beta_N XY - \mu_{1N}X \\
\frac{\mathrm{d}U}{\mathrm{d}t'} &= \hat{\beta}_N XY + r_N Y - (\tau_N + \mu_{2N})U \\
\frac{\mathrm{d}Y}{\mathrm{d}t'} &= \hat{\tau}_N U - \gamma_N YZ - \mu_{3N}Y \\
\frac{\mathrm{d}Z}{\mathrm{d}t'} &= \nu_N Y - \mu_{4N}Z,
\end{aligned} \tag{3.7}$$

where

$$\begin{aligned}
&\lambda_N = \frac{\lambda t_0}{x_0}, \quad \beta_N = \beta y_0 t_0, \quad \hat{\beta}_N = \frac{\beta x_0 y_0 t_0}{u_0}, \quad r_N = \frac{r y_0 t_0}{u_0}, \quad \tau_N = \tau t_0, \\
&\hat{\tau}_N = \frac{\tau u_0 t_0}{y_0}, \quad \gamma_N = \gamma z_0 t_0, \quad \nu_N = \frac{\nu y_0 t_0}{z_0}, \quad \mu_{1N} = \mu_1 t_0, \quad \mu_{2N} = \mu_2 t_0, \\
&\mu_{3N} = \mu_3 t_0, \quad \mu_{4N} = \mu_4 t_0.
\end{aligned}$$

The subscript N above designates the various quantities as being associated with the non-dimensionalised system. Note that $\hat{\beta}_N = \frac{x_0}{u_0}\beta_N$ and $\hat{\tau}_N = \frac{u_0}{y_0}\tau_N$. The corresponding non-dimensional feasible region is

$$\Gamma_N := \left\{ (X,U,Y,Z) \in \mathbb{R}^4_+ : \right.$$
$$\left. X \le \frac{\lambda_N}{\mu_{1N}},\ X + \frac{u_0}{x_0}U + \frac{y_0}{x_0}Y \le \frac{\lambda_N}{\mu_N - \frac{r_N u_0}{y_0}},\ Z \le \frac{\nu_N \lambda_N}{\mu_{4N}\left(\mu_N - \frac{r_N u_0}{y_0}\right)} \right\},$$

where $\mu_N = \min\{\mu_{1N}, \mu_{2N}, \mu_{3N}\}$. Using the same methods as before, which are explained in detail in Appendix A.1, we can derive the expression for the non-dimensional basic reproduction number for viral infection for the non-dimensionalised model (3.7) to be

$$R_{0N} = \frac{\hat{\tau}_N}{\mu_{3N}(\tau_N + \mu_{2N})}\left(\hat{\beta}_N X_H + r_N\right), \quad \text{where } X_H = \frac{\lambda_N}{\mu_{1N}}. \tag{3.8}$$

3.5.2 Re-scaling Variables to Order 1

In this section, we evaluate the relative sizes of each cell population based on rough estimates from experimental data on T-cell dynamics and infection by HTLV-I, and select appropriate factors x_0, u_0, y_0, z_0 to re-scale all variables to order 1. Note that this approach is slightly different to the usual method of non-dimensionalisation which reduces the number of parameters but does not necessarily re-scale all variables to order 1. Our approach, which in general introduces new parameters to the model, allows us to consider the specific immunological properties of HTLV-I infection represented by our mathematical model in the context of biologically realistic parameter values, of which some but not all have been

clinically measured. It can be expected that even known parameters will vary over a broad range of values. Nevertheless, we can exploit those that are known by restricting their selection to reasonable, averaged values. Meanwhile, quantities that are not yet known are initially chosen to be biologically plausible, and changes to these values are explored further to better understand their role in describing infection using our mathematical model. Our non-dimensional re-scaling has been chosen according to HTLV-I infection in ACs, as this group comprises the vast majority of HTLV-I-infected individuals (Bangham, 2000a; Proietti et al., 2005).

The key to selecting an appropriate non-dimensional re-scaling is to take a closer look at the average T-cell populations during the course of chronic infection. First, since the normal $CD4^+$ helper T-cell count in a healthy individual averages around 850 cells/mm^3, this suggests an appropriate scaling for the X-equation to be $x_0 = 1000$ cells/mm^3. Next, it has been estimated that approximately 1% or less of an AC's $CD4^+$ helper T-cells are provirus-positive, and that out of these infected target cells, only 1% to 5% are activated and expressing viral antigens at any given time (Asquith and Bangham, 2008; Asquith et al., 2005b; Richardson et al., 1997). We therefore select scalings for the latently infected (the U-equation) and actively infected (the Y-equation) target cell populations to be $u_0 = 10$ cells/mm^3 and $y_0 = 0.5$ cells/mm^3, respectively. Meanwhile, the normal level of $CD8^+$ cytotoxic T-lymphocytes typically lies between 300 cells/mm^3 and 800 cells/mm^3 of peripheral blood (Bofill et al., 1992). It has been estimated that up to 1% or more of circulating $CD8^+$ CTLs can recognise HTLV-I epitopes (Bangham, 2000b). Since chronic HTLV-I infection is often accompanied by high frequencies of HTLV-I-specific $CD8^+$ CTLs (Asquith et al., 2005a; Bangham and Osame, 2005), we calculate, using the upper limit of $CD8^+$ CTL frequency in the peripheral blood, that $z_0 = 1$ $CD8^+$ cell/mm^3 is a good approximation for the case of CTLs specific to HTLV-I. Lastly, since the timescale during the acute phase of infection is not known experimentally, we non-dimensionalise time by setting $t_0 = 1$ day. We assume that this choice of non-dimensionalisation re-scales each of the variables X, U, Y, Z to order 1. The corresponding non-dimensional parameters are shown in Table 3.2.

3.5.3 Re-scaling Parameter Values to Order 1

With the choice of non-dimensional re-scaling of variables to order 1 and the chosen parameter values as in Table 3.2, we are now in a position to evaluate the relative impact of each term in the equations to the solution of the non-dimensionalised model (3.7). It is seen that the highest order parameter in each of the equations is 10^{-2}. This suggests another re-scaling, this time for the non-dimensional parameters, to make them of order one.

Non-dimensional Scaling	Value
x_0	1000 cells/mm^3
u_0	10 cells/mm^3
y_0	0.5 cells/mm^3
z_0	1 cells/mm^3
t_0	1 day
Non-dimensional Initial Condition	**Value**
$X(0)$	1
$U(0)$	0.01
$Y(0)$	0.05
$Z(0)$	0.01
Non-dimensional Parameter	**Value**
$\lambda_N = \frac{\lambda t_0}{x_0}$	0.010
$\beta_N = \beta y_0 t_0$	0.0005
$\hat{\beta}_N = \frac{\beta x_0 y_0 t_0}{u_0}$	0.05
$r_N = \frac{r y_0 t_0}{u_0}$	0.00055
$\tau_N = \tau t_0$	0.003
$\hat{\tau}_N = \frac{\tau u_0 t_0}{y_0}$	0.06
$\gamma_N = \gamma z_0 t_0$	0.029
$\nu_N = \frac{\nu y_0 t_0}{z_0}$	0.018
$\mu_{1N} = \mu_1 t_0$	0.012
$\mu_{2N} = \mu_2 t_0$	0.03
$\mu_{3N} = \mu_3 t_0$	0.03
$\mu_{4N} = \mu_4 t_0$	0.03

Table 3.2: Table of non-dimensional parameter values for the non-dimensionalised model (3.7).

Let $\epsilon = u_0/x_0 = 10^{-2}$. The corresponding non-dimensional parameters, re-scaled to order 1 and written in terms of the un-scaled non-dimensional and dimensional parameters for reference, are shown in Table 3.3. Note that in Table 3.3, the parameter $\tau_R = 0.3$ is not re-scaled to order 1 as it is seen to always go together with the parameter $\mu_{2R} = 3$, which is of order 1. By re-scaling most of the parameters of the non-dimensional model (3.7) to order 1, we obtain the following equivalent system:

$$\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t'} &= \epsilon(\lambda_R - \epsilon\beta_R XY - \mu_{1R}X) \\
\frac{\mathrm{d}U}{\mathrm{d}t'} &= \epsilon(\hat{\beta}_R XY + \epsilon r_R Y - (\tau_R + \mu_{2R})U) \\
\frac{\mathrm{d}Y}{\mathrm{d}t'} &= \epsilon(\hat{\tau}_R U - \gamma_R YZ - \mu_{3R}Y) \\
\frac{\mathrm{d}Z}{\mathrm{d}t'} &= \epsilon(\nu_R Y - \mu_{4R}Z).
\end{aligned} \tag{3.9}$$

On this timescale, the dynamics occur at a very slow rate as all of the equations are

Re-scaled, Non-dimensional Parameter	Value
$\lambda_R = \frac{\lambda_N t_0}{\epsilon} = \frac{\lambda t_0}{\epsilon x_0}$	1
$\beta_R = \frac{\beta_N t_0}{\epsilon^2} = \frac{\beta y_0 t_0}{\epsilon^2}$	5
$\hat{\beta}_R = \frac{\hat{\beta}_N t_0}{\epsilon} = \frac{\beta x_0 y_0 t_0}{\epsilon u_0}$	5
$r_R = \frac{r_N t_0}{\epsilon^2} = \frac{r y_0 t_0}{\epsilon^2 u_0}$	5.5
$\tau_R = \frac{\tau_N t_0}{\epsilon} = \frac{\tau t_0}{\epsilon}$	0.3
$\hat{\tau}_R = \frac{\hat{\tau}_N t_0}{\epsilon} = \frac{\tau u_0 t_0}{\epsilon y_0}$	6
$\gamma_R = \frac{\gamma_N t_0}{\epsilon} = \frac{\gamma z_0 t_0}{\epsilon}$	2.9
$\nu_R = \frac{\nu_N t_0}{\epsilon} = \frac{\nu y_0 t_0}{\epsilon z_0}$	1.8
$\mu_{1R} = \frac{\mu_{1N} t_0}{\epsilon} = \frac{\mu_1 t_0}{\epsilon}$	1.2
$\mu_{2R} = \frac{\mu_{2N} t_0}{\epsilon} = \frac{\mu_2 t_0}{\epsilon}$	3
$\mu_{3R} = \frac{\mu_{3N} t_0}{\epsilon} = \frac{\mu_3 t_0}{\epsilon}$	3
$\mu_{4R} = \frac{\mu_{4N} t_0}{\epsilon} = \frac{\mu_4 t_0}{\epsilon}$	3

Table 3.3: Table of non-dimensional parameter values, re-scaled to order 1, for the re-scaled, non-dimensionalised models (3.9) and (3.10).

changing on the order of $\epsilon = 10^{-2}$. Indeed, this observation echoes our remark on the time series plot of the original dimensional model (3.1) (i.e. the 'baseline model') that motivated our analysis in the first place, where the behaviour of the solution was seen to vary on the timescale of $1/\epsilon = 100$ days (refer to Figure 3.2). This suggests a re-scaling of time to $T = \epsilon t'$ so that

$$\frac{\mathrm{d}}{\mathrm{d}T} = \frac{\mathrm{d}t'}{\mathrm{d}T}\frac{\mathrm{d}}{\mathrm{d}t'} = \frac{1}{\epsilon}\frac{\mathrm{d}}{\mathrm{d}t'}.$$

Noting that $\hat{\beta}_R = \beta_R$ and $\hat{\tau}_R = 20\tau_R$, the model now becomes

$$\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}T} &= \lambda_R - \epsilon\beta_R XY - \mu_{1R}X \\
\frac{\mathrm{d}U}{\mathrm{d}T} &= \beta_R XY + \epsilon r_R Y - (\tau_R + \mu_{2R})U \\
\frac{\mathrm{d}Y}{\mathrm{d}T} &= 20\tau_R U - \gamma_R YZ - \mu_{3R}Y \\
\frac{\mathrm{d}Z}{\mathrm{d}T} &= \nu_R Y - \mu_{4R}Z.
\end{aligned} \tag{3.10}$$

Assuming that all the re-scaling and parameter choices are biologically realistic, we observe, by comparing the terms in each equation, that

(i) $\epsilon\beta_R \ll \lambda_R, \mu_{1R}$ in the X-equation, and

(ii) $\epsilon r_R \ll \beta_R, \tau_R + \mu_{2R}$ in the U-equation.

Hence, the impact of these two quantities on the overall dynamics of the solution is 'small' in relation to the other terms, at least with this choice of non-dimensional re-scaling and

parameter values. We will denote the corresponding feasible region for model (3.10) as Γ_R. The re-scaled, non-dimensional basic reproduction number for viral infection is

$$R_{0R} = \frac{20\tau_R}{\mu_{3R}(\tau_R + \mu_{2R})}\Big(\beta_R X_H + \epsilon r_R\Big), \quad \text{where } X_H = \frac{\lambda_R}{\mu_{1R}}. \tag{3.11}$$

The choice of re-scaled, non-dimensional parameter values as in Table 3.3 represents biologically realistic quantities that arise during chronic HTLV-I infection and correspond to the situation when the basic reproduction number, R_{0R}, takes a value greater than unity. By a result analogous to Theorem 3.2 for the re-scaled, non-dimensional model (3.10), the (re-scaled, non-dimensionalised) infection-free equilibrium P_{0R} is unstable and the existence and the global stability of a unique chronic infection equilibrium P_R^* in $\mathring{\Gamma}_R$ is guaranteed. Choosing biologically realistic initial conditions, we plot the time evolution graph for the re-scaled, non-dimensional model (3.10) in Figure 3.3. We point out several observations about Figure 3.3. First, we notice that the population of healthy target cells, $X(T)$, begins of order 1 and remains so as time progresses. Meanwhile, the populations of the infected target cells, $U(T)$ and $Y(T)$, and immune cells, $Z(T)$, begin of much smaller order, $O(\epsilon)$, and appear to increase exponentially before approaching asymptotically to the chronic infection equilibrium levels. In the following section, we perform an asymptotic expansion to characterise the transient behaviour of solutions of the re-scaled, non-dimensionalised model (3.10).

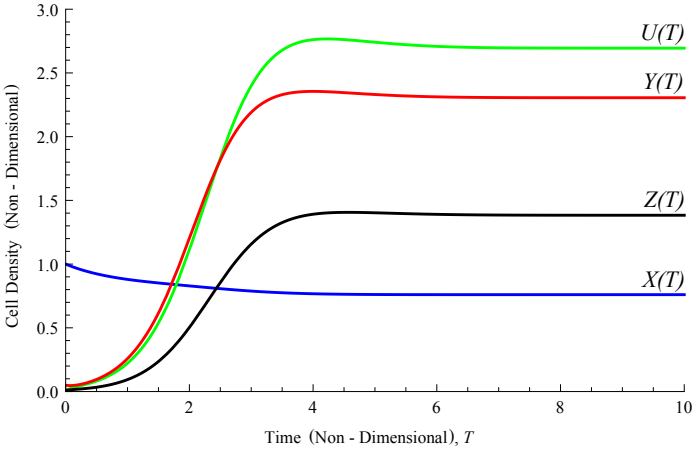


Figure 3.3: Convergence to P_R^* when $R_{0R} > 1$ for the re-scaled, non-dimensional model described by system (3.10). Initial conditions are biologically plausible and, in this case, are chosen to be $X(0) = 1, U(0) = 0.01, Y(0) = 0.05, Z(0) = 0.01$. The plot shows that after the initial behaviour, all populations of target cells and immune cells persist at an equilibrium level. Given the parameter values from Table 3.3, $R_0 \approx 2.56 > 1$.

3.5.4 Asymptotic Analysis

In this subsection, we study the transient and asymptotic behaviour of model (3.10) when $R_{0R} > 1$, for biologically realistic initial conditions. Recall that for model (3.10), we selected appropriate re-scalings so that all variables and parameter values may be assumed to be of order 1 in the chronic phase of infection. The size of each compartment, however, is bound to differ at the onset of the infection. For example, the initial level of viraemia exposed to the host is likely to be significantly lower than the level during chronic infection after the virus has replicated and established itself in the host. Likewise, due to the functionality of adaptive immunity, the reaction of the host immune response towards the virus is often substantially less vigorous at the start of the infection (i.e. when the immune system is naïve against the new invading pathogen) than later on as time progresses and virus-specific immunity has been built up. For these reasons, let us assume that the initial level of infected target cells, as well as the initial level of HTLV-I-specific $CD8^+$ CTLs, are small and, in particular, of order ϵ. Meanwhile, we assume that at the beginning of the infection, nearly all target cells are uninfected and thus the level of healthy cells is of order 1.

Let us look at the initial behaviour in more detail. Specifically, we examine the exponential transient growth of the U, Y, Z-compartments of model (3.10) during the initial phase of infection when the variables $U(T), Y(T), Z(T) \sim O(\epsilon)$, and search for an explicit analytical expression characterising this initial exponential growth. By plotting the initial transients alongside the equilibrium solutions, we are able to get an idea as to how the two piece together smoothly to form the solution of model (3.10).

This amounts to examining the behaviour of model (3.10) with initial conditions $X(0) = O(1)$, and $U(0), Y(0), Z(0) = O(\epsilon)$. Under these initial conditions, several terms in the equations for model (3.10) contribute very little to the dynamics of each T-cell population under investigation, at least at the onset of the infection, and can therefore be omitted when examining the transient behaviour of solutions. For instance, in the X-equation, the non-linear term $\epsilon\beta_R XY$ is initally of order ϵ^2 as $Y(0) \sim O(\epsilon)$, and thus may be discarded as the other terms are of order 1. Similarly, the U, Y, Z-equations are all initially of order ϵ, hence the $\epsilon r_R Y$ term in the U-equation and the non-linear term $\gamma_R YZ$ in the Y-equation are both of order ϵ^2 and may also be discarded. Note that the $\beta_R XY$ term in the U-equation cannot be discarded as it is of order ϵ, i.e. the same order as the initial value $Y(0)$. As our primary interest here is in the initial behaviour, let us introduce the notation $\tilde{X}, \tilde{U}, \tilde{Y}$ to denote the corresponding transient variables of the (non-dimensionalised, re-scaled) healthy, latently infected, and actively infected target cell populations. Similarly, denote by $\tilde{Z}$ the transient variable for the (non-dimensionalised, re-scaled) compartment of HTLV-I-specific

CTLs. Specifically, $\tilde{X}$, $\tilde{U}$, $\tilde{Y}$, and $\tilde{Z}$ are solutions to the linearisation of model (3.10) about the initial conditions $\tilde{X}(0) = O(1)$, and $\tilde{U}(0)$, $\tilde{Y}(0)$, $\tilde{Z}(0) = O(\epsilon)$, namely,

$$\begin{aligned}
\frac{\mathrm{d}\tilde{X}}{\mathrm{d}T} &= \lambda_R - \mu_{1R}\tilde{X} \\
\frac{\mathrm{d}\tilde{U}}{\mathrm{d}T} &= \beta_R(\tilde{X} - X_0)Y_0 + \beta_R X_0 \tilde{Y} - (\tau_R + \mu_{2R})\tilde{U} \\
\frac{\mathrm{d}\tilde{Y}}{\mathrm{d}T} &= 20\tau_R \tilde{U} - \mu_{3R}\tilde{Y} \\
\frac{\mathrm{d}\tilde{Z}}{\mathrm{d}T} &= \nu_R \tilde{Y} - \mu_{4R}\tilde{Z},
\end{aligned} \tag{3.12}$$

where $X_0 = \tilde{X}(0)$. As a result of the linearisation, the $\tilde{X}$-equation is decoupled from the remaining equations and can be solved directly to yield

$$\tilde{X}(T) = X_H + (X_0 - X_H)e^{-\mu_1 T}. \tag{3.13}$$

We know that $X_H \sim O(1)$ due to our choice of non-dimensional re-scaling, hence from the above formula with initial value $X_0 \sim O(1)$, we see that the variable $\tilde{X}(T)$ starts out to be of $O(1)$ and remains $O(1)$ as time progresses, changing slowly as it eventually converges to X_H. Note that this behaviour is slightly different to that in the non-dimensionalised, re-scaled model (3.10), where the non-linear term $\epsilon\beta_R XY$ in the X-equation plays a role in determining the long-term behaviour of $X(T)$ by applying a damping effect on $X(T)$ so that it converges not to X_H, but instead to the chronic infection equilibrium level $\bar{X}$ which, although not the same as X_H, is nevertheless still of order 1.

Next, we examine the transient behaviour of the remaining coupled equations. As the $\tilde{X}$ variable in the linearised non-dimensionalised, re-scaled model (3.12) varies little in magnitude and remains of order 1 throughout its time course according to Equation (3.13), we make the assumption that $\tilde{X}(T)$ in the $\tilde{U}$-equation can be described by its initial constant value X_0 so that

$$\frac{d\tilde{U}}{dT} = \beta_R X_0 \tilde{Y} - (\tau + \mu_2)\tilde{U}.$$

As we will see shortly, by choosing X_0 to be $\bar{X}$, the value at the chronic infection equilibrium, we find that the initial transient behaviour matches quite closely to the solution of system (3.10) with the same initial condition. The transient variables $\tilde{U}, \tilde{Y}, \tilde{Z}$ are solutions to the linear subsystem

$$\begin{bmatrix} d\tilde{U}/dT \\ d\tilde{Y}/dT \\ d\tilde{Z}/dT \end{bmatrix} = \underbrace{\begin{bmatrix} -\tau_R - \mu_{2R} & \beta_R X_0 & 0 \\ 20\tau_R & -\mu_{3R} & 0 \\ 0 & \nu_R & -\mu_{4R} \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} \tilde{U} \\ \tilde{Y} \\ \tilde{Z} \end{bmatrix}. \tag{3.14}$$

We can solve this linear first-order system with constant coefficient matrix $\mathcal{A}$ by finding the fundamental matrix $e^{\mathcal{A}T}$. We begin by determining the eigenvalues of the matrix $\mathcal{A}$. The characteristic polynomial is

$$\chi_R(\zeta) = (\zeta + \mu_{4R})\Big[\zeta^2 + (\tau_R + \mu_{2R} + \mu_{3R})\zeta - \mu_{3R}(\tau_R + \mu_{2R})\big[g(X_0) - 1\big]\Big],$$

where

$$g(X_0) = \frac{20\beta_R\tau_R X_0}{\mu_{3R}(\tau_R + \mu_{2R})}.$$

Notice that $g(X_0)$ looks like the expression for R_0 in the case when $\epsilon = 0$, but with the initial condition X_0 in place of the level of healthy target cells at the infection-free steady state X_H. In the absence of infection, it is reasonable to presume that $X_0 = X_H$, or at least $|X_0 - X_H| \ll O(1)$, so that $g(X_0) \approx R_0$. Since our current focus is to study the transient dynamics leading to chronic infection, we may assume that $g(X_0) > 1$. Thus, there are three distinct, real eigenvalues of the constant coefficient matrix $\mathcal{A}$, one of which is positive. These are precisely the roots of $\chi_R(\zeta) = 0$: $\zeta_1 = \phi_+ > 0$, $\zeta_2 = \phi_- < 0$, and $\zeta_3 = -\mu_{4R}$, where

$$\phi_\pm = -\frac{1}{2}(\tau_R + \mu_{2R} + \mu_{3R}) \pm \frac{1}{2}\sqrt{(\tau_R + \mu_{2R} + \mu_{3R})^2 + 4\mu_{3R}(\tau_R + \mu_{2R})\big[g(X_0) - 1\big]}.$$

Standard calculations from linear algebra then yield three linearly independent eigenvectors $\vec{p}_1$, $\vec{p}_2$, and $\vec{p}_3$, each corresponding to its respective eigenvalue ζ_1, ζ_2, and ζ_3:

$$\vec{p}_1 = \begin{bmatrix} \frac{1}{20\tau_R}(\phi_+ + \mu_{3R}) \\ 1 \\ \frac{\nu_R}{\phi_+ + \mu_{4R}} \end{bmatrix}, \quad \vec{p}_2 = \begin{bmatrix} \frac{1}{20\tau_R}(\phi_- + \mu_{3R}) \\ 1 \\ \frac{\nu}{\phi_- + \mu_{4R}} \end{bmatrix} \quad \text{and} \quad \vec{p}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, the explicit form of the transient solution is

$$\begin{bmatrix} \tilde{U}(T) \\ \tilde{Y}(T) \\ \tilde{Z}(T) \end{bmatrix} = \vec{q}_1 e^{\phi_+ T} + \vec{q}_2 e^{\phi_- T} + \vec{q}_3 e^{-\mu_{4R} T}, \tag{3.15}$$

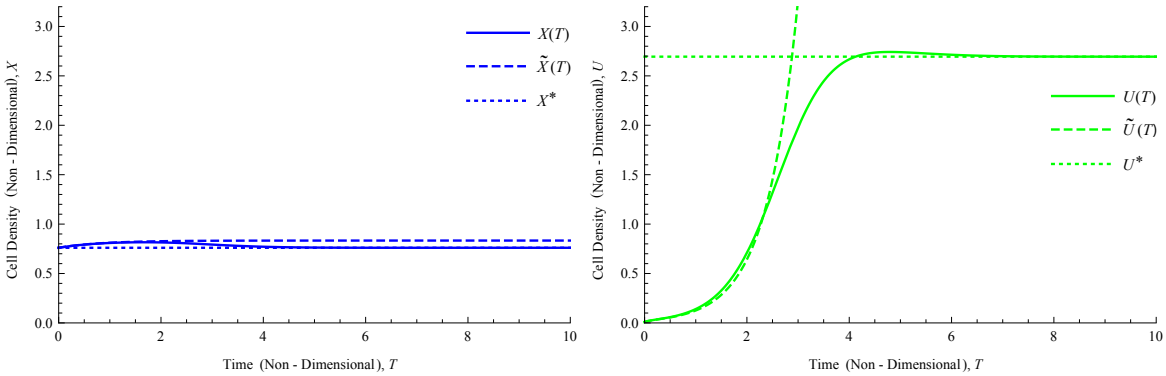
where $\vec{q}_1$, $\vec{q}_2$, $\vec{q}_3$ are derived from the eigenvectors and depend on the initial conditions X_0, U_0, Y_0, and Z_0. The formula given by Equation (3.15) represents an explicit solution of the linearised subsystem (3.14) and accurately describes the solution of the full re-scaled, non-dimensional model (3.10) in the initial stages of the infection until $T \approx 1/\phi_+$, at which point the transient solution becomes of order 1 and non-linear effects start to come into play.

By taking a closer look at the initial transient behaviour when $R_0 > 1$ as in Figure 3.3, we see that at the very start, the solutions $U(T)$, $Y(T)$, $Z(T)$ of model (3.10) are relatively flat.

After a short time has progressed, the variables begin to increase exponentially. This sharp increase corresponds to the positive eigenvalue $\zeta_1 = \phi_+$ becoming dominant in the transient variables $\tilde{U}(T), \tilde{Y}(T), \tilde{Z}(T)$. We can derive a crude upper bound for the size of ϕ_+ by noting that for any two positive real numbers $a_1, a_2 \in \mathbb{R}_+$, the inequality $\sqrt{a_1 + a_2} \leq \sqrt{a_1} + \sqrt{a_2}$ holds. In particular, applying this inequality yields the following estimate:

$$\phi_+ \leq \sqrt{\mu_{3R}(\tau_R + \mu_{2R})\big[g(X_0) - 1\big]}. \tag{3.16}$$

For each of the variables, we can visualise the way in which the initial transients $\tilde{X}(T)$, $\tilde{U}(T)$, $\tilde{Y}(T)$, $\tilde{Z}(T)$, which are solutions of the linearised, re-scaled, non-dimensional model (3.12), piece together with the solution at the chronic infection equilibrium $\bar{X}$, $\bar{U}$, $\bar{Y}$, $\bar{Z}$ by plotting the two alongside the solution $X(T)$, $U(T)$, $Y(T)$, $Z(T)$ of the non-dimensionalised, re-scaled model (3.10), as in Figures 3.4(a)–(d).



(a) Healthy target cells: $X(T)$, $\tilde{X}(T)$, X^*.

(b) Latently infected target cells: $U(T)$, $\tilde{U}(T)$, U^*.

(c) Actively infected target cells: $Y(T)$, $\tilde{Y}(T)$, Y^*.

(d) HTLV-I-specific CTLs: $Z(T)$, $\tilde{Z}(T)$, Z^*.

Figure 3.4: Piecing together transient solutions with the chronic infection equilibrium P_R^*. Transient solutions of system (3.12) are shown as dashed lines, the equilibrium solutions at P_R^* are shown as dotted lines, and the solutions of the original non-dimensionalised model (3.10) are shown as solid lines.

The results of non-dimensionalisation and asymptotic analysis performed in this section has shown two main features of the baseline model (3.1) under our assumptions on the parameter values and re-scalings, which have been based on biologically realistic quantities. First, mitotic transmission does not appear to play a significant role in driving the dynamics of the system, although it is important in determining the levels of each compartment at equilibrium. Second, the linearisation of the re-scaled, non-dimensional model is able to accurately describe the behaviour of solutions to the full non-linear system, at least during the initial transient stages of infection. Whereas we cannot determine an explicit solution for the original re-scaled, non-dimensional model (3.10), we can, by contrast, write an explicit formula for the transient solution of the linearised version of the model given by system (3.12).

3.6 Measures of Viral Burden

One of the most important tasks faced by clinicians after identifying an HTLV-I seropositive individual is to evaluate the severity of the infection. How much damage has the virus caused to the host? Is the infected individual asymptomatic or are there signs of chronic or malignant disease? For the former, is the individual at risk of developing symptoms? For the latter, how far and how fast has the disease progressed? The degree of detriment caused by the virus is termed the *viral burden*.

The most common measure of viral burden in the host is the proviral load, and it is given by the proportion of all target cells in the peripheral blood that carry a provirus, regardless of their state of activation (Asquith et al., 2005a; Asquith and Bangham, 2008). With the advance of current experimental techniques, such a measure, which simply reports the magnitude of the virus infection, is relatively easy to obtain from host sera and is frequently reported in experimental data as a percentage of all peripheral blood mononuclear cells (PBMCs) (Asquith et al., 2005a). In terms of our mathematical model, we define the proviral load at equilibrium as the infected fraction of all $CD4^+$ helper T-cells,

$$v^* = \frac{u^* + y^*}{x^* + u^* + y^*}. \qquad \Big[\text{Proviral Load (PVL)}\Big] \tag{3.17}$$

A defining characteristic of our mathematical modelling approach to HTLV-I infection is the separation of proviral cells into two distinct compartments based on infected target cell latency or activation. This allows us to consider how the two populations of proviral cells co-exist and how they persist as the infection propagates back and forth through the latent and activated states. In our model, the total number of infected target cells, both latent and active, at equilibrium is given by the expression $\left(u^* + y^*\right)$. Thus, we can define and

examine a second measure of viral burden to be the proportion of proviral cells that are activated, which is represented by the following ratio:

$$\frac{y^*}{u^* + y^*}. \qquad \Big[\text{Active Proportion of Proviral Cells}\Big] \tag{3.18}$$

3.7 Exploring Spontaneous Viral Expression: Why is HTLV-I not silent? (Question 1.)

In HTLV-I infection, the vast majority of proviral cells do not express viral proteins at any given time, an observation that lends support to the traditional viewpoint that HTLV-I is largely inactive (Asquith and Bangham, 2008; Asquith et al., 2007). Nevertheless, it has been shown that most HTLV-I seropositive individuals, regardless of clinical status, mount a vigorous virus-specific cellular immune response, suggesting continuous stimulation by expressed viral antigens and thereby raising the important question (Asquith et al., 2000): 'Is HTLV-I infection really silent?' This issue has been resolved by Asquith et al. (2007), who have demonstrated experimentally that each day, a small proportion of infected target cells spontaneously re-activate and express viral antigens. Such studies have been crucial in establishing that the strategy of HTLV-I infection is not that of complete latency, but one that is more dynamic.

In light of our overall theme of understanding HTLV-I persistence, in this section, we consider the following conundrum: Why is HTLV-I not silent? Indeed, the expression of viral proteins accompanying infected T-cell activation exposes proviral cells to immune surveillance, which are then at risk of destruction. If this is the case, then why would the evolution of HTLV-I favour a route for viral expression and activation rather than remaining completely latent? The fact is that viral activation can be both beneficial and detrimental to the proviral cell. On the one hand, displaying viral proteins is required for infectious cell-to-cell transmission and is believed to drive the rapid, selective expansion of infected cells (Asquith and Bangham, 2008; Igakura et al., 2003). On the other hand, it simultaneously allows such cells to be eliminated by a persistently activated HTLV-I-specific CTL response (Asquith et al., 2005a; Bangham, 2003). The balance between the two opposing selection forces determines the outcome of infection. To resolve our conundrum, we switch our perspective to that of the virus to understand why and how viral activation can be advantageous.

Our mathematical model (3.1) explicitly incorporates the interactions among the latent and active states of HTLV-I proviral cells, and virus-specific CTLs. In particular, the transition from the latently infected target cell compartment to the actively infected target cell compartment is encapsulated in the parameter τ, the rate of spontaneous viral expression

that accompanies proviral activation. We can therefore examine the impact of the parameter τ on the long-term behaviour of model (3.1). Our results are consistent with those of Li and Lim (2011), whose model explored in detail the role of Tax expression on chronic HTLV-I infection, but in the absence of a dynamically changing host immune response.

3.7.1 Tax/HBZ Expression Drives Chronic Infection and is an Important Determinant of Proviral Load

The basic reproduction number for viral infection R_0, defined in Equation (3.2), can be thought of as a measure of viral success with respect to the ability to establish and propagate the infection. A simple calculation shows that R_0 is an increasing function of τ; that is, the rate of spontaneous viral activation is seen to be a factor that drives the system towards chronic infection, making it difficult for infected individuals to clear the virus.

Next, we explore the way in which the expression of viral proteins affects the proviral load. In Figure 3.5, we have plotted numerically a curve from our mathematical model (3.1) showing the impact of the rate τ of spontaneous infected target cell activation on the equilibrium proviral load v^*. Our model results show that for small values of τ, there is a sharp positive, rather than negative, correlation between τ and v^*. As the value of τ is increased, we observe that the magnitude of the proviral load v^* begins to level off before gradually decreasing and settling to a constant positive level for sufficiently large values of τ. Hence, viral expression is a factor that determines an infected individual's equilibrium proviral load. This second point provides further support for a non-latent virus strategy.

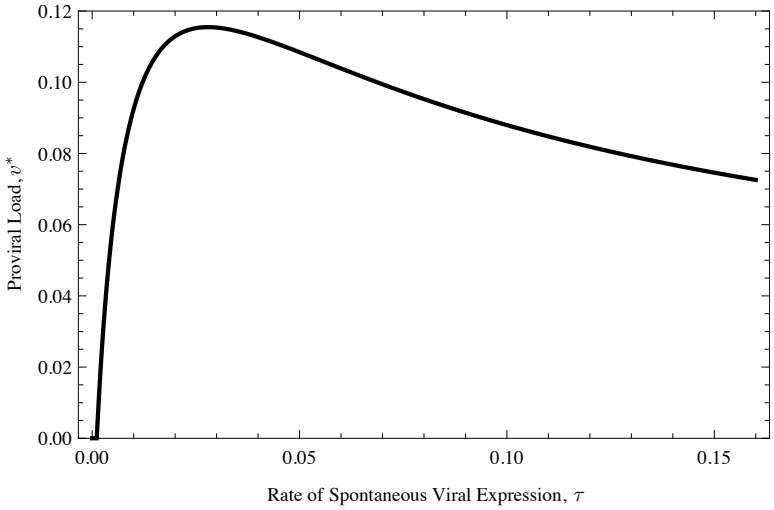


Figure 3.5: The effect of spontaneous viral activation, τ, and subsequent expression of viral antigens on the equilibrium proviral load, v^*, as governed by our mathematical model (3.1). The above graph was obtained by solving the equilibrium equations for model (3.1) numerically using the parameter values from Table 3.1 to determine x^*, u^* and y^*, then plotting the quantity $v^* = \big(u^* + y^*\big)/\big(x^* + u^* + y^*\big)$ as a function of the parameter τ.

A traditional measure of viral success is the proviral load, and a higher proviral load is a better established infection. With this perspective, Figure 3.5 can be thought of as reflecting a cost-benefit relationship for infected target cell activation that governs successful viral persistence: A rate of spontaneous viral expression that is too low is of little benefit to the virus as latency does not propagate the spread of infection; meanwhile, one that is too high could also be detrimental as it over-exposes proviral cells to host immunity and places a stronger demand on the virus to replicate fast enough to outpace CTL-mediated lysis. Our results suggest that an intermediate rate of spontaneous infected T-cell activation, for example in the vicinity of the peak of the graph in Figure 3.5, could be optimal for HTLV-I to successfully establish and maintain a high proviral load in the presence of host immunity. This observation supports the theoretical hypothesis by Asquith and Bangham (2008) that HTLV-I relies on a dynamic balance between viral latency and activation in order to persist indefinitely in the host: complete transcriptional latency is counter-productive to viral transmission and propagation of the infection, whereas full infected T-cell activation leaves too many proviral cells susceptible to immune-mediated lysis. Nevertheless, there may be another consequence of a higher rate of spontaneous viral protein expression despite its association with lowering the magnitude of the proviral load, which comes to light by examining the activated fraction of infected cells.

3.7.2 Tax/HBZ Expression Increases Activated Fraction of Infected Cells

Model (3.1) allows us to consider the effect of spontaneous viral expression or activation, represented by the parameter τ, on the second measure of viral burden or detriment, the equilibrium proportion of proviral cells that are actively infected, $y^*/\left(u^* + y^*\right)$. The result of numerical investigation of this dependence is shown in Figure 3.6, where we observe that the net effect of a faster rate of spontaneous viral expression is a higher active proportion. In addition, unlike the proviral load v^*, which decreases when τ is sufficiently large, the active proportion of proviral cells is a monotonically increasing function of τ. The main implication of the result here is that the rate of spontaneous viral expression, τ, determines a characteristic ratio of latent versus activated infected cells at equilibrium, irrespective of the magnitude of the proviral load itself.

Our model predicts a relationship between the proviral load v^* and the active proportion of infected cells $y^*/\left(u^* + y^*\right)$ which is consistent with the recent experimental results of Melamed et al. (2013), who demonstrated a negative correlation between the abundance of a given infected T-cell clone and the proportion of the respective clone that spontaneously expresses the provirus in HAM/TSP patients. Taken together, the results shown in Figures 3.5 and 3.6 imply such a negative correlation, when the rate of spontaneous viral

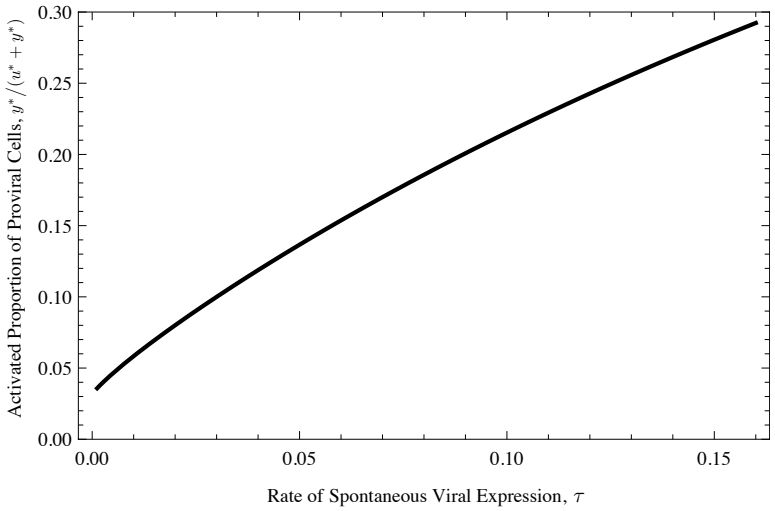


Figure 3.6: A faster rate of proviral expression τ increases the proportion of infected target cells that are transcriptionally active, $y^*/(u^*+y^*)$, at equilibrium. The above graph was obtained by determining the steady state solutions during chronic infection for model (3.1) numerically, then plotting the quantity $y^*/(u^*+y^*)$ as a function of the parameter τ. All parameter values have been selected from Table 3.1.

expression τ exceeds a threshold value (of about 0.03 day^{-1} as in Figure 3.5, i.e. past the peak; see also Figure 3.11).

3.8 Evaluating CTL Quality: What determines the strength of the HTLV-I-specific CTL response, and why is infection life-long? (Question 2.)

In the previous section, we investigated the impact of spontaneous viral expression in our model to try and resolve the issue as to why HTLV-I is not silent. We have discovered that a low level of viral protein expression can offer tremendous benefits in terms of viral propagation and success. In other words, the net effect of viral activation is to drive chronic infection and aid the persistence of HTLV-I. These observations, in turn, suggest a complex interaction with the host immune response that we now explore in further detail.

3.8.1 CTL Frequency $\neq$ CTL Efficiency

It is well established that cellular immunity, driven by the selective expansion of antiviral CTLs, is an important branch of the human immune system that aids in eliminating invading intracellular pathogens such as viruses (Parham, 2005). In the case of HTLV-I, virus-specific CTLs have been shown experimentally to be highly efficient at killing antigen-expressing proviral cells and are an important determinant of an infected individual's proviral load (Asquith et al., 2005a). Moreover, CTLs are thought to play a role in the devel-

opment of HTLV-I-associated pathologies (Asquith and Bangham, 2000; Bangham et al., 1999; Jeffery et al., 1999), although there is conflicting speculation as to whether CTLs are protective or whether they contribute to disease progression (Biddison et al., 1997; Mosley and Bangham, 2009). It is therefore essential to investigate how CTLs affect the outcome of chronic HTLV-I infection.

The strength of the immune response against virally infected cells has traditionally been thought of as being directly related to the frequency or magnitude of virus-specific CTLs in the peripheral blood. In HTLV-I infection, large numbers of circulating anti-HTLV-I CTLs are often found in blood samples from HTLV-I seropositive individuals, whether or not malignant disease is present (Bangham, 2003; Biddison et al., 1997; Parker et al., 1992). Such an observation is somewhat unexpected as the vast majority of the proviral load is latent and therefore invisible to host immunity. Indeed, how is it possible for the CTL response to have any impact on the course of infection when most proviral cells are transcriptionally silent at any given time and are thus not even visible to immune surveillance? Asquith and Bangham (2008) hypothesise that since the expression of HTLV-I proteins is crucial for viral propagation, then by exerting selective pressure on the small proportion of virus-expressing cells, CTLs block a key point in the virus life cycle and can therefore have a substantial effect on the infection dynamics.

This underlies an important issue that arises in experimental data on HTLV-I-infected individuals, which is identifying the aspects of cellular immunity that are responsible for controlling the proviral load during the chronic phase of infection, otherwise known as the quality of the CTL response. A natural question comes to mind: What determines CTL quality? Morever, is it the size of the CTL response that influences its ability to reduce the proviral load, or is there another factor at play?

We again turn to our baseline mathematical model (3.1) for some insights, focussing on two specific factors commonly believed to impact the proviral load of an infected individual: CTL frequency and CTL-mediated lysis of antigen-expressing infected cells. In our model, the equilibrium proviral load is given by v^*. Moreover, the abundance of CTLs is precisely z^* and the rate at which CTLs eliminate activated proviral cells is represented by the parameter γ. A few straightforward calculations in our model yield Theorem 3.3, which demonstrates that the proviral load is positively correlated with the level of CTLs at equilibrium and negatively correlated with the rate of CTL-mediated lysis.

Theorem 3.3. *Assume that $R_0 > 1$. Then, the equilibrium proviral load v^* is an increasing function of z^*, but a decreasing function of γ.*

Proof of Theorem 3.3. Consider the proviral load at equilibrium, which is given by the expression $v^* = \big(u^* + y^*\big)\big/\big(x^* + u^* + y^*\big)$. From Equation (3.4), we have

$$u^* + y^* = \frac{\mu_4}{\tau\nu}(\gamma z^* + \tau + \mu_3)z^* \quad \text{and} \quad x^* = \frac{\lambda\nu}{\beta\mu_4 z^* + \nu\mu_1}.$$

To show the first part of the theorem, we compute

$$\frac{\mathrm{d}}{\mathrm{d}z^*}(u^* + y^*) = \frac{\mu_4}{\tau\nu}(2\gamma z^* + \tau + \mu_3) > 0,$$

and

$$\frac{\mathrm{d}x^*}{\mathrm{d}z^*} = -\frac{\beta\lambda\nu\mu_4}{(\beta\mu_4 z^* + \nu\mu_1)^2} < 0.$$

Then,

$$\begin{aligned}
\frac{\mathrm{d}v^*}{\mathrm{d}z^*} &= \frac{\mathrm{d}}{\mathrm{d}z^*}\left(\frac{u^* + y^*}{x^* + u^* + y^*}\right) \\
&= \frac{(x^* + u^* + y^*)\frac{\mathrm{d}}{\mathrm{d}z^*}(u^* + y^*) - (u^* + y^*)\frac{\mathrm{d}}{\mathrm{d}z^*}(x^* + u^* + y^*)}{(x^* + u^* + y^*)^2} \\
&= \frac{1}{(x^* + u^* + y^*)^2}\left(x^*\frac{\mathrm{d}}{\mathrm{d}z^*}(u^* + y^*) - (u^* + y^*)\frac{\mathrm{d}x^*}{\mathrm{d}z^*}\right) \\
&= \frac{\mu_4}{\tau\nu(x^* + u^* + y^*)^2}\left[(2\gamma z^* + \tau + \mu_3)x^* + \frac{\beta\lambda\tau\nu^2(u^* + y^*)}{(\beta\mu_4 z^* + \nu\mu_1)^2}\right] \\
&> 0.
\end{aligned}$$

Hence, the proviral load at equilibrium, v^*, is an increasing function of z^*.

The second part of the theorem is proved by observing that

$$\frac{\mathrm{d}v^*}{\mathrm{d}\gamma} = \frac{\mathrm{d}v^*}{\mathrm{d}z^*}\frac{\mathrm{d}z^*}{\mathrm{d}\gamma}.$$

It is straight-forward to show that $\mathrm{d}z^*\big/\mathrm{d}\gamma < 0$ (the derivation follows directly from differentiating the equation $F(z^*) = 0$, where $F(z)$ is defined in Equation (3.5), implicitly with respect to the parameter γ). Combined with the result from the first part of the theorem demonstrating that $\mathrm{d}v^*\big/\mathrm{d}z^* > 0$, we conclude that

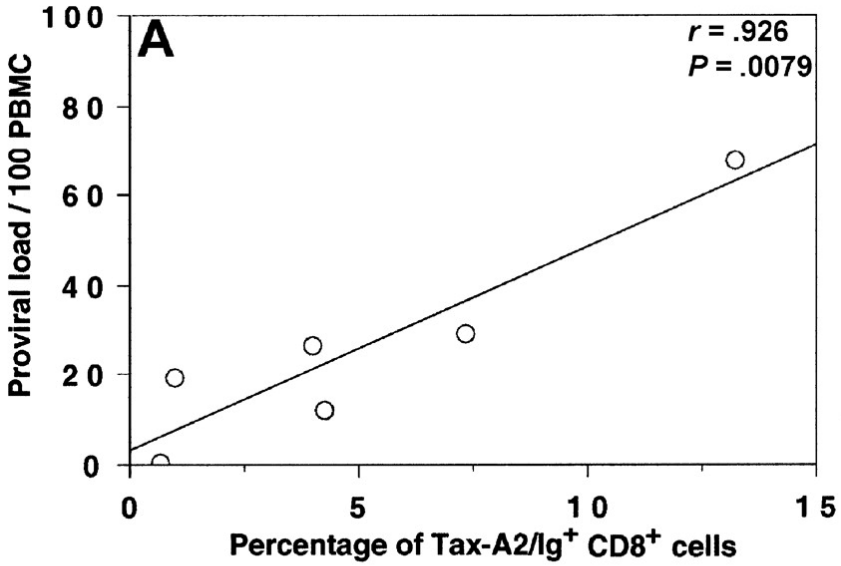
$$\frac{\mathrm{d}v^*}{\mathrm{d}\gamma} < 0;$$

that is, v^* is a decreasing function of γ. □

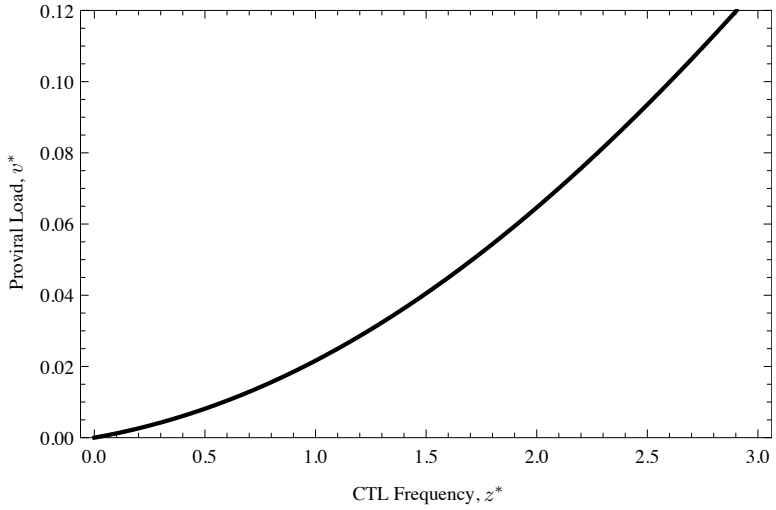
What the result of Theorem 3.3 suggests is that the strength or efficacy of the CTL response with respect to controlling the proviral load is influenced by how fast the HTLV-I-specific CTL response can eliminate virus-infected target cells rather than the size of the CTL pool. A significant implication of this result is that the traditional line of thinking that the strength of the anti-HTLV-I CTL response is equal to its size needs to be re-evaluated. Indeed, intuitively one may expect that the more CTLs there are, the stronger the immune response, and thus the lower the proviral load. However, a decrease in the proviral load subsequently implies reduced stimulation for the proliferation of CTLs that control the proviral load. Such opposing feedback makes it unclear how one can predict the relationship between CTL frequency and the strength of the CTL response with respect to its effect on the proviral load, a sentiment which has been expressed by experimentalists and theoreticians (Asquith and Bangham, 2008; Bangham et al., 2009; Elemans et al., 2012). Nevertheless, the notion that large numbers of CTLs can be associated with large numbers of virus-infected cells is not new, and has been observed in clinical data (Kubota et al., 2000, 2003; Nagai et al., 2001; Wodarz et al., 2001). Moreover, experiments have demonstrated that a fast rate of clearance of virus-expressing infected cells is associated with a reduced proviral load (Asquith et al., 2005a; Kattan et al., 2009).

Figure 3.7 shows a positive correlation between the frequency of HTLV-I-specific CTLs and the proviral load in an HTLV-I seropositive individual. In Figure 3.7(a), we reproduce the experimental results of Nagai et al. (2001), who measured the percentage of CTLs targeting the immunodominant viral Tax protein in the context of HLA-A2, which represented the bulk of the anti-HTLV-I CTL response, and the viral burden was measured in terms of the number of proviral cells per 100 peripheral blood mononuclear cells (PBMCs), which is directly proportional to the number of $CD4^+$ helper T-cells that are infected. Meanwhile, Figure 3.7(b) plots the relationship between the number of CTLs at equilibrium z^* and the proviral load in terms of the infected fraction of target cells v^* as determined by our baseline mathematical model (3.1). Although quantitative comparisons have not been done, our theoretical results agree qualitatively with the experimental data from Nagai et al. (2001).

Figure 3.8 displays a negative correlation between the rate at which actively infected target cells expressing viral antigens are cleared by $CD8^+$ CTLs, and the proviral load. Asquith et al. (2005a) reported their experimental measurements of anti-viral efficacy as the number of Tax^+ cells cleared per day per CTL and its correlation with the proviral load in terms of PBMCs (Figure 3.8(a)). As Tax is the immunodominant viral antigen targeted by CTLs, such a measure is directly related to the parameter γ in our model. In addition, the viral burden measured in terms of infected PBMCs corresponds directly to the measure of the proviral load with respect to the $CD4^+$ helper T-cell population, which is represented in



(a) Experimental data. Reproduced from Nagai et al. (2001) (Figure 6A) with permission from the publisher.

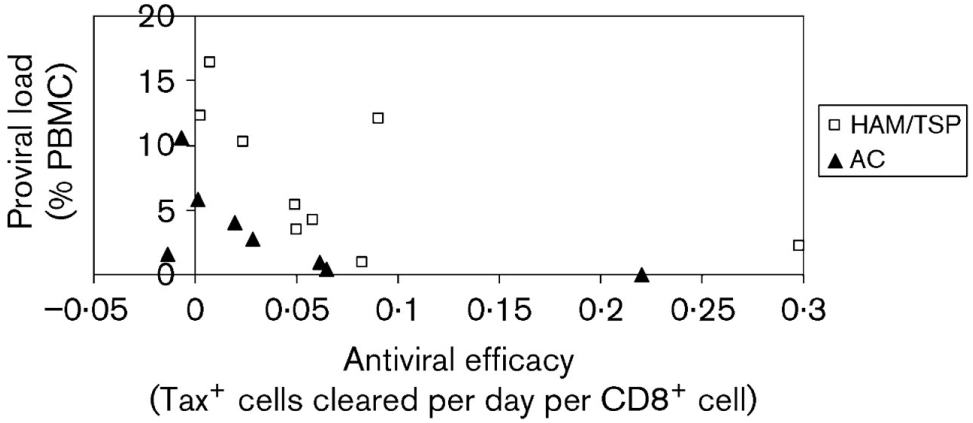


(b) Model results.

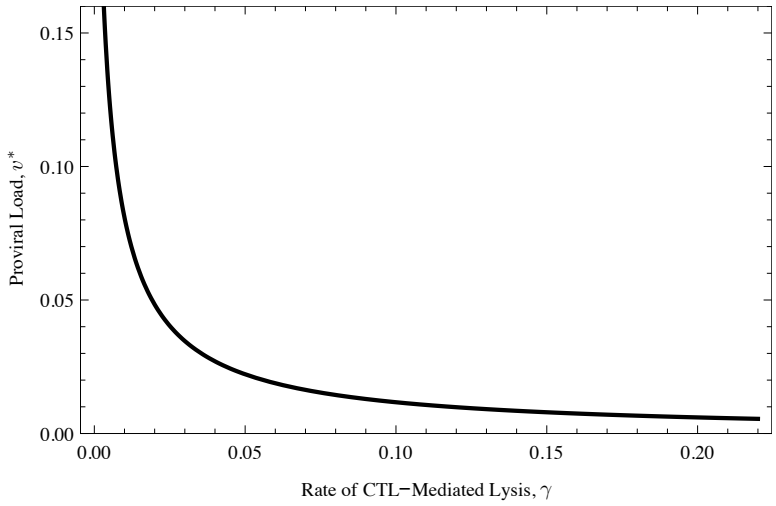
Figure 3.7: Positive correlation between CTL frequency (horizontal axis) and proviral load (vertical axis). (a) The experimental results from Nagai et al. (2001) show data points and have been fitted theoretically with a linear curve. (Figure reproduced with permission from the publisher.) (b) The corresponding curve from our model (3.1), which plots the relationship between the proviral load v^* and CTL abundance z^* at equilibrium, displays a different form, but agrees qualitatively with the experimental data. Although there is a difference in the numerical values along the corresponding axes between the two graphs, the measures of the proviral load for each graph are proportional to one other. However, the exact scaling is unknown. Parameter values have been selected as in Table 3.1.

our model by the quantity v^*. We observe that the data from Asquith et al. (2005a) are in qualitative agreement with Figure 3.8(b), which shows how the equilibrium proviral load v^*

changes with respect to γ according to our baseline mathematical model (3.1). Our model



(a) Experimental data. Reproduced from Asquith et al. (2005a) (Figure 4) with permission from the publisher.



(b) Model results.

Figure 3.8: Negative correlation between rate of CTL-mediated lysis (horizontal axis) and the proviral load (vertical axis). (a) The experimental measurements from Asquith et al. (2005a) show data points for both ACs and HAM/TSP patients. (Figure reproduced with permission from the publisher.) (b) Our model results arising from the system of equations (3.1) are in qualitative agreement. Parameter values are taken from Table 3.1.

therefore supports the relatively recent idea that a large CTL response is not necessarily a strong one, and that CTL frequency is not a reliable measure of immune efficiency (Asquith and Bangham, 2008; Bangham et al., 2009; Bangham, 2009; Bangham et al., 1996, 1999; Elemans et al., 2012). In addition, our findings agree with previous mathematical modelling work by Nowak and Bangham (1996) and Wodarz et al. (2001) demonstrating how the CTL response to persistent viral infections such as HIV and HTLV-I can be inefficient at controlling the proviral load yet CTLs may exist at high frequencies.

3.8.2 Efficient CTLs Cannot Eliminate the Active Proportion of Proviral Cells

We have shown that effective immune control of the HTLV-I proviral load is reflected in the rate at which CTLs eliminate actively infected cells rather than the size of the CTL pool. This finding motivates the question: Why is there no cure for HTLV-I infection? Indeed, our results indicate that a fast rate of CTL-mediated lysis is effective at reducing the proviral load. Yet, the application of vaccines that boost the efficiency of CTL responses elicited by infected hosts has been unsuccessful at achieving complete elimination of the virus, and infection remains chronic (Bangham, 2000a).

Using our baseline mathematical model (3.1), we offer a possible explanation for this failure by examining the impact of CTL efficiency, encapsulated by the parameter γ, on the proportion of proviral cells that is activated, $y^*/(u^* + y^*)$. We observe from numerical investigations that for sufficiently large values of γ, further changes in γ have little effect on the active proportion of proviral cells (note that we can equally conclude from the results shown in Figure 3.8(b) that, when this same value of γ is exceeded, there is little further reduction in the proviral load v^*). Indeed, combining this result with our earlier finding that the rate of spontaneous viral expression τ displays a positive correlation with the active proportion of proviral cells yields Figure 3.9, which illustrates that the quantity $y^*/(u^*+y^*)$ steadily increases with larger values of τ for fixed values of γ, whereas for fixed τ, an increase in γ has only a minimal impact on $y^*/(u^* + y^*)$ when γ is sufficiently large.

The main implications of the above result are that: (1) an efficient CTL response fails to eradicate the infection because it cannot eliminate the activated (i.e. 'aggressive') part of the infected target cell population, and (2) a faster rate of proviral activation which exposes more infected cells to immune surveillance makes it more difficult for CTLs to achieve clearance of activated proviral cells. These observations motivate a hypothesis relating to pathogenesis which is discussed in the following section.

3.9 Insights to HTLV-I-Associated Disease: What characterises HTLV-I-linked pathology? (Question 3.)

A puzzling issue in HTLV-I infection is why only a small fraction of chronically infected individuals develop the inflammatory disease HAM/TSP whereas the vast majority remain as life-long asymptomatic carriers (ACs) of the virus. Although much experimental work has shed light on the way in which several key factors of the infection may be associated with clinical status, either AC or HAM/TSP patient, a clear determinant to distinguish between the two states is not yet known.

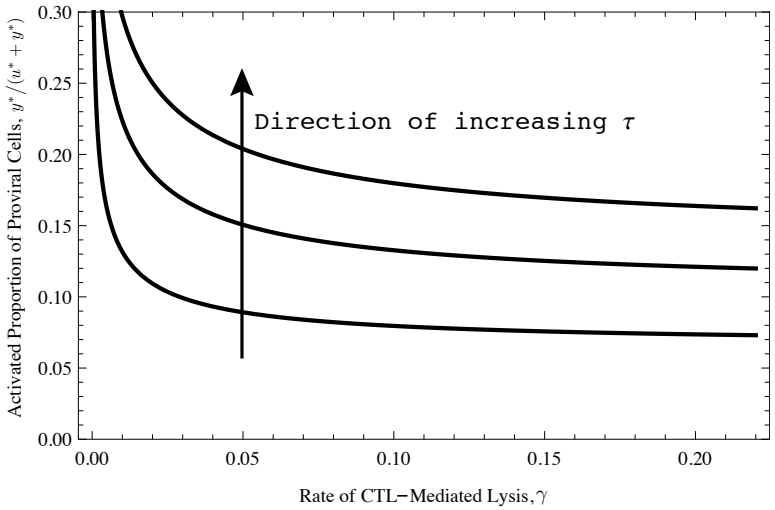


Figure 3.9: Evaluating the effects of CTL-mediated lysis, γ, and spontaneous viral expression, τ, on the activated proportion of cells, $y^*/(u^*+y^*)$. The above figure depicts a bifurcation diagram for $y^*/(u^*+y^*)$ with respect to γ for three different choices of τ, and was obtained by solving model (3.1) at equilibrium numerically. CTL quality or effectiveness may be characterised by high values of γ. For fixed values of τ, when γ is sufficiently large, a further increase in γ has little effect on the $\text{Tax}^+/\text{HBZ}^+$ proportion of proviral cells $y^*/(u^*+y^*)$. Therefore, even though efficient CTLs (i.e. large γ) reduce the magnitude of the proviral load, they do not necessarily decrease the active (i.e. 'aggressive') proportion of proviral cells. Here, we have selected values of τ to be 0.03, 0.07 and 0.11 day^{-1}. All other parameters have been chosen as in Table 3.1.

In the previous sections, our investigations of key factors including spontaneous activation of proviral cells, CTL frequency, and CTL efficiency have illuminated important properties of HTLV-I infection that help our understanding of viral persistence in the presence of chronically activated immune responses. In this section, we take our results one step further by interpreting the roles of these key parameters in the context of HTLV-I-associated disease. In particular, we propose a plausible hypothesis for the identification of disease status and suggest a possible mechanism for pathogenesis.

3.9.1 A Hypothesis for Determining Clinical Status

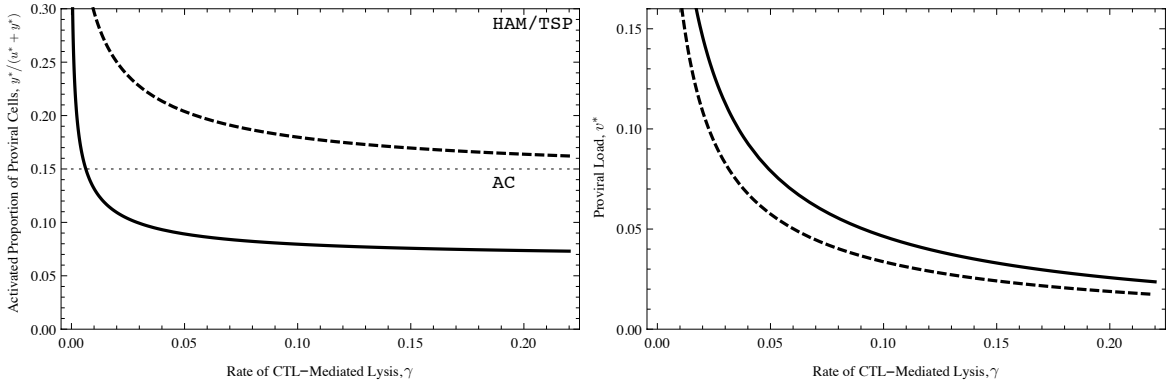
The equilibrium proviral load is a major correlate of clinical status: a high proviral load is associated with HAM/TSP development, whilst a low proviral load is associated with asymptomatic carriage. However, it is known that the size of the proviral load does not, on its own, separate ACs from HAM/TSP patients. Indeed, there is broad overlap among the proviral loads of ACs and HAM/TSP patients, and a high proviral load is not a reliable determinant of disease status (Asquith and Bangham, 2007; Asquith et al., 2005b).

We propose instead that the extent of activation of proviral cells, $y^*/(u^*+y^*)$, and

not necessarily the absolute magnitude of the proviral load, v^*, may be associated with the inflammatory disease, HAM/TSP. For example, there may be a threshold level for the activated proportion of infected target cells above which asymptomatic carriers develop disease, regardless of the size of the proviral load. At this stage, it is important to clarify that such a proposed hypothesis of disease status is not due to the measurement of the proportion of activated cells per se, but because the proportion of activated cells is a direct reflection of the activation parameter τ, representing the rate of spontaneous viral expression, with which it is positively correlated. This activation parameter τ represents a characteristic property of the virus that can be thought of as a measure of how "aggressive" it is. A more aggressive virus will have a faster activation rate, and a faster activation rate has been shown experimentally to be associated with pathology; namely, it has been demonstrated that infected cells from HAM/TSP patients expressed the Tax protein at higher levels than those from ACs (c.f. Figure 2 from (Asquith et al., 2005b)).

This hypothesis for distinguishing between ACs and HAM/TSP patients is shown graphically in Figure 3.10(a), where the proposed theoretical threshold for pathogenesis is represented by the thin, horizontal dotted line. With respect to our interpretation, the threshold for pathogenesis in terms of the activated proportion of proviral cells as shown is determined wholly by the rate of spontaneous viral expression τ for a given individual with all other parameters fixed. Our hypothesis is consistent with previous experimental data showing that infected cells from HAM/TSP patients expressed the Tax protein at higher levels than those from ACs (c.f. Figure 2 from Asquith et al., 2005b).

In chronic HTLV-I infection, it is well-documented that the magnitude of the proviral load is not a direct reflection of the extent of activation of proviral cells. Although generally a high proviral load is associated with disease progression, it is not uncommon for asymptomatic carriers to harbour proviral loads that are higher than those of symptomatic carriers (Asquith et al., 2005b). The separation of clinical status based on the active proportion of infected target cells rather than the magnitude of the proviral load could potentially explain how infected individuals with a low proviral load could have HAM/TSP due to a high proportion of activated infected cells, or conversely, how those with a high proviral load could be asymptomatic due to a low proportion of activated infected cells. For a wide range of parameter values, our model exhibits behaviour that is consistent with experimental observations (c.f. Figure 2(b) from Asquith and Bangham, 2008); that is, at a given rate of CTL lysis, the proviral load in ACs is lower than that in HAM/TSP patients (results not shown). However, due to the shape of the curve describing how the activated proportion of infected cells changes with the proviral load and because the proposed theoretical threshold for disease manifestation is not precisely known, our model can



(a) The thin horizontal dotted line indicates a theoretical threshold for disease manifestation.

(b) A high proviral load is neither necessary nor sufficient to identify the presence of disease.

Figure 3.10: A hypothesis for classifying clinical status arising from our mathematical model (3.1) is that a high activated proportion of infected cells, and not necessarily the size of the proviral load, may be an important determinant of disease. In the above figures, we solve model (3.1) at equilibrium numerically and plot the bifurcation diagrams for the activated proportion of cells, $y^*/(u^*+y^*)$, with respect to the rate of CTL-mediated lysis, γ, for two different choices of the rate of spontaneous viral expression, τ. The solid line and the dashed line correspond, respectively, to two different individuals. (a) For all values of γ, the individual represented by the solid line is seen to have a lower activated proportion of infected cells than that of the individual represented by the dashed line. (b) By contrast, the proviral load for the former individual (solid line) is, in fact, higher than that of the latter individual (dashed line) for all values of γ. Furthermore, for sufficiently high rates of CTL-mediated lysis γ, the individual corresponding to the solid line has an active proportion of infected cells that lies below the theoretical threshold for disease manifestation (the thin, dotted, horizontal line) and would therefore be classified as an asymptomatic carrier under our hypothesis (refer again to (a)), despite having a higher absolute magnitude of the proviral load than the individual corresponding to the dashed line (refer again to (b)). Here, we have selected values of τ to be 0.03 day^{-1} (solid line) and 0.11 day^{-1} (dashed line). All other parameters have been chosen as in Table 3.1.

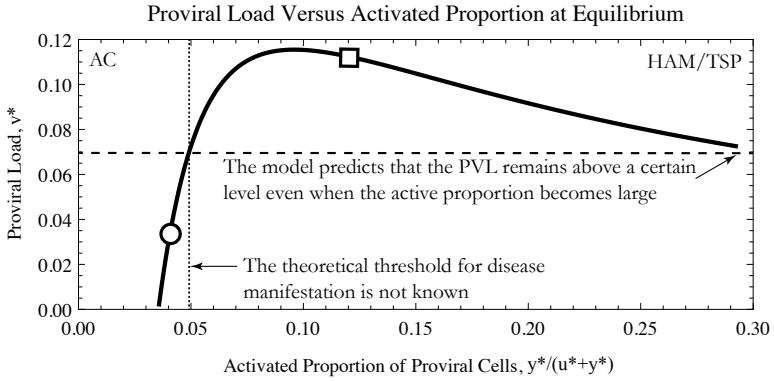
also exhibit the converse phenomenon in which the proviral load in an AC is greater than that of a HAM/TSP patient (for more details, we refer the reader to Figures 3.5 and 3.6, as well as Figure 3.11). Indeed, such a scenario is possible in our model as highlighted in Figure 3.10, where we have plotted the efficiency of the CTL response γ against the two different measures of viral detriment. Parameter values are selected that lie in the same range as those in Table 3.1. In Figure 3.10(a), with respect to the activated proportion of proviral cells, the solid line lies below the threshold for pathogenesis for all sufficiently large values of γ, and thus corresponds to asymptomatic clinical status, while the dashed line lies above the proposed threshold for all γ and would therefore represent a patient with HAM/TSP. Yet with respect to the proviral load v^* as in Figure 3.10(b), it is seen that for all values of γ, the magnitude of v^* for the AC (given by the solid line) is, in fact, higher

than that for the HAM/TSP patient (given by the dashed line). An alternative way to visualise the complex connections between the proviral load and clinical status with respect to the activated proportion of proviral cells, according to our hypothesis, is illustrated in Figure 3.11, which shows two theoretically possible scenarios that can be captured using the baseline model (3.1), and lends support to the experimental observation that the magnitude of the proviral load is neither necessary nor sufficient to classify clinical status. Our model therefore allows for a consistent framework that could potentially account for this peculiar aspect of chronic HTLV-I infection which has been well-documented in the literature (Asquith et al., 2005b; Bangham et al., 1999; Nagai et al., 1998). Moreover, given two infected individuals whose immune efficiencies are equal, we remark that our model can demonstrate the intuitive scenario whereby the proviral load of an AC is lower than that of a HAM/TSP patient, as shown in Figure 3.11(a). Nevertheless, further studies are needed to deduce better estimates for the position of the theoretical threshold for disease manifestation and determine whether the alternative, less intuitive scenario is possible, namely, can there exist two HTLV-I seropositive individuals with the same immune efficiency, one AC and one HAM/TSP patient, in which the proviral load of the AC is higher than that of the HAM/TSP patient, as shown in Figure 3.11(b)?

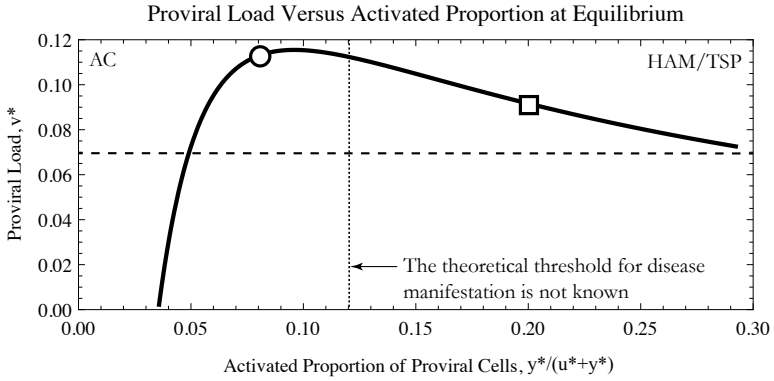
3.9.2 Failure of HAM/TSP Treatment

We have observed previously how increasing the rate of CTL-mediated lysis γ, which is a measure of CTL efficiency, is effective in lowering the magnitude of the proviral load, but is ineffective in reducing the activated proportion of infected target cells. In terms of the efficiency of the CTL response, there is an additional consequence of our hypothesis for identifying clinical status. Taking another look at Figure 3.10(a), we notice that it is possible for chronically infected individuals to have less efficient CTLs (i.e. a low value of γ) and still be asymptomatic because of having an active proportion of infected cells that lies below the designated threshold. Alternatively, it may also be the case that chronically infected individuals with very efficient CTLs (i.e. a high value of γ) could be diagnosed with HAM/TSP due to an active proportion of infected cells that lies above the proposed threshold.

The above observation has an implication for HAM/TSP treatments and could offer a possible explanation as to why a cure for the disease has been difficult to discover. Consider an infected individual with HAM/TSP who has a given immune efficiency γ, as indicated by the open circle $\bigcirc$ in Figure 3.12. The dashed line shows how the activated proportion of proviral cells of this individual (i.e. with all other parameters fixed) varies with different rates of CTL-mediated lysis. Suppose that treatment is administered to enhance the efficiency



(a) Case 1. The proviral load of an AC is always lower than that of a HAM/TSP patient.



(b) Case 2. Our model results demonstrate that it is possible for the proviral load of an AC to be higher than that of a HAM/TSP patient.

Figure 3.11: Bifurcation diagram relating proviral load, v^*, and the active proportion of infected cells, $y^*/(u^* + y^*)$. The above figures show the outcomes of solving model (3.1) at equilibrium numerically with the rate of spontaneous viral expression, τ, as the bifurcation parameter, and plotting the corresponding relationship between v^* and $y^*/(u^* + y^*)$. The results of our model are consistent with experimental and clinical observations, and can even account for some of the peculiarities of HTLV-I infection. In particular, it is known that although the magnitude of the proviral load of an HTLV-I seropositive individual is correlated with the presence of disease, there exists a broad overlap in the magnitudes of the proviral loads among ACs and HAM/TSP patients, and a high proviral load is neither necessary nor sufficient for the progression of HTLV-I-associated disease. Here, the rate γ of CTL-mediated lysis is fixed. The thin, dotted vertical line representing the theoretical threshold for disease manifestation corresponds to the thin, dotted horizontal line in Figure 3.10(a). The precise position of this theoretical threshold is unknown. (a) If the threshold for disease manifestation is sufficiently low, then the proviral load of an AC (along the curve to the left of the threshold, for example, at the position of the open circle ◯) would always be lower than that of a HAM/TSP patient (along the curve to the right of the threshold, for example, at the position of the open square □). This represents the typical case whereby a HAM/TSP patient harbours a higher proviral load than that of an AC. (b) However, if the threshold for disease manifestation is sufficiently high, then our model demonstrates the theoretical possibility for the proviral load of an AC (along the curve to the left of the threshold, for example, symbolised by an open circle ◯) to be above that of a HAM/TSP patient (along the curve to the right of the threshold, for example, symbolised by an open square □). All parameters have been selected from Table 3.1.

of the HTLV-I-specific CTL response which, although lowering the absolute magnitude of the proviral load v^*, does not significantly impact the proportion of proviral cells that are actively infected (c.f. Figure 3.10, dashed lines). The active proportion continues to lie above the threshold for pathogenesis and the individual remains with the inflammatory disease.

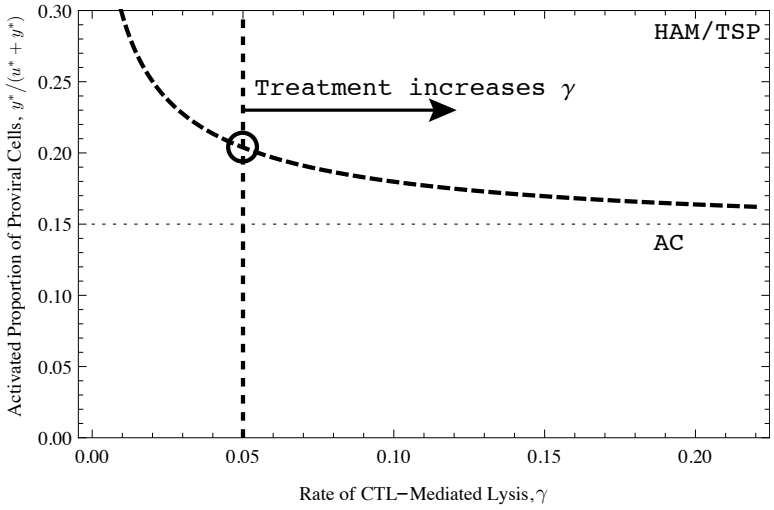


Figure 3.12: Failure of HAM/TSP treatment. In the above figure, we solve model (3.1) at equilibrium numerically, and plot the bifurcation diagram for the activated proportion of cells, $y^*/(u^*+y^*)$, with respect to the rate of CTL-mediated lysis, γ, for a fixed rate of spontaneous viral expression, τ. Our results suggest a plausible reason as to why treatment for HTLV-I pathology is unsuccessful. Consider the profile of an infected individual characterised by a given initial rate of CTL-mediated lysis (thick, vertical, dotted line) and a corresponding activated proportion of proviral cells which, before treatment is administered, is indicated by the open circle ◯ in the figure. The dashed line shows the way that the activated proportion varies with the rate of CTL-mediated lysis according to model (3.1). Although the administration of vaccines that improve the efficiency of anti-HTLV-I CTLs reduce the magnitude of the proviral load (as shown in Figure 3.10(b), dashed line), it may be ineffective at lowering the activated proportion of proviral cells below the theoretical threshold for pathogenesis (thin, horizontal, dotted line). Here, we have selected $\tau = 0.11$ day^{-1}. All other parameters have been chosen as in Table 3.1.

3.9.3 Immune Compromise is a Possible, Though Unlikely, Cause of Disease

An intriguing aspect of chronic HTLV-I infection is that only a small fraction of infected individuals develop the inflammatory disease, HAM/TSP, after a long asymptomatic clinical phase. It remains unresolved as to how or why disease may manifest over time during the chronic stage of infection. In the context of our hypothesis, if there does exist a threshold level for the activated proportion of infected target cells that distinguishes clinical status,

how could it be possible for the active proportion to increase and eventually cross the proposed threshold?

One possible route, which occurs in HIV infection, is that the virus severely compromises the human immune system, and a weakened state allows for the progression of disease. For instance, over time HTLV-I may disrupt the functionality of the immune system and render virus-specific CTLs ineffective at controlling viral replication, allowing the population of proviral cells to expand tremendously. The scenario is illustrated in Figure 3.13, where we first consider an asymptomatic carrier with a given rate of CTL-mediated lysis γ, represented by the open circle ◯. The solid black line shows how the active proportion of proviral cells varies with respect to immune efficiency γ and all other parameters fixed. It is reasonable to believe that as time progresses, the efficiency of the CTL response is compromised so that the value of γ becomes very small. In this case, the active proportion of the viral burden would increase and could end up above the threshold for pathogenesis, which is shown in Figure 3.13 as an open square □. The infected individual then develops HAM/TSP. However, it has been experimentally shown that chronic infection by HTLV-I is not associated with overt immuno-suppression (Asquith and Bangham, 2007); more or less normal T-cell functionality remains. If anything, the efficiency of a functional CTL response is thought to increase over time, rather than decrease, due to continued exposure and increased specificity towards target antigens (Parham, 2005). This explanation is therefore unlikely to describe the progression of HAM/TSP.

3.9.4 A Plausible Mechanism for Pathogenesis

Although silencing the expression of viral proteins allows HTLV-I to evade immune selection, there is little benefit for proviral cells to be fully transcriptionally latent. Indeed, continued propagation of the infection requires that at least a small fraction of latently infected cells becomes spontaneously re-activated each day. As we have seen previously in Figure 3.6, the rate of spontaneous infected T-cell activation, τ, displays a strict positive correlation with the active proportion of proviral cells, $y^*/(u^* + y^*)$. In terms of the development of the inflammatory disease, HAM/TSP, we offer a hypothesis that as time progresses, the active proportion of proviral cells may increase, and could eventually cross the threshold for pathogenesis, thus causing disease. One way that this could happen is that the rate τ of spontaneous expression of viral proteins, which in our model is treated as a constant, may in fact increase over time.

A hypothetical situation is illustrated in Figure 3.14. As before, we consider an asymptomatic carrier with a specified CTL efficiency γ. The position of this individual is given by the open circle ◯ on the graph, and the solid black curve on which the open circle lies

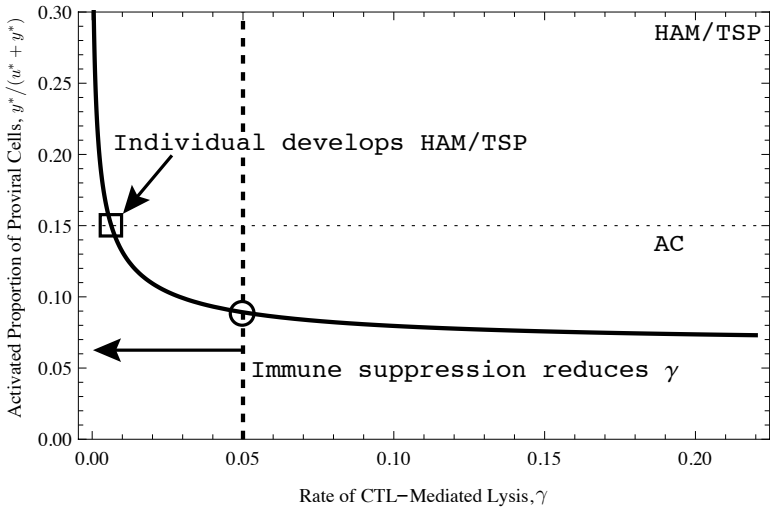


Figure 3.13: A prediction of the baseline model (3.1) is that immune compromise leading to development of disease is possible. The above figure shows the outcome of solving model (3.1) at equilibrium numerically, and plotting the bifurcation diagram for the activated proportion of cells, $y^*/(u^*+y^*)$, with respect to the rate of CTL-mediated lysis, γ, for a fixed rate of spontaneous viral expression, τ. Consider an infected individual whose immune system exhibits a given efficiency in terms of the rate γ of CTL-mediated lysis (thick, vertical, dotted line), classifying him or her as being initially an AC and is represented in the figure by the open circle ◯. The solid black curve illustrates how the active proportion of proviral cells, $y^*/(u^*+y^*)$, for this particular individual changes with respect to the efficiency of his or her anti-viral CTL response according to model (3.1). Suppose that the immune system of this individual is compromised, resulting in a decrease in CTL efficiency which, in turn, leads to an increase in the active proportion (moving left along the solid black line). At some sufficiently low value of γ, the active proportion rises above the threshold for pathogenesis (thin, horizontal, dotted line), proposing a route by which the infected individual could develop disease, such as the neurological and inflammatory condition HAM/TSP. Disease manifestation is represented by the open square □ in the figure. However, chronic infection by HTLV-I is not associated with immune suppression, suggesting that such an explanation for disease progression is unlikely. Here, we have selected $\tau = 0.03$ day^{-1}. All other parameters have been chosen as in Table 3.1.

is the way that the active proportion of proviral cells changes with respect to the rate of CTL-mediated lysis, γ, with all other parameters fixed. Suppose that over time, the rate of spontaneous infected target cell activation τ increases. The net effect, as expected, would be a general increase in the active proportion of infected cells. It is then possible that the proportion of actively infected cells reaches the threshold for pathogenesis, represented by the open square □ in Figure 3.14, and manifestation of the inflammatory disease, HAM/TSP, occurs.

However, it is known that the vast majority of chronically infected individuals do not develop HAM/TSP, and instead remain asymptomatic throughout their lifetime. Suppos-

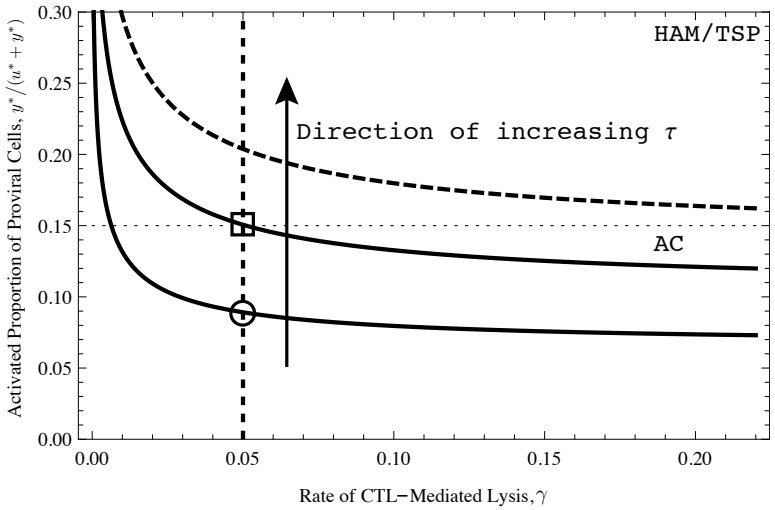


Figure 3.14: Our baseline mathematical model (3.1) suggests a plausible mechanism for the development of disease. The above figure illustrates the outcomes of solving model (3.1) at equilibrium numerically, and plotting the bifurcation diagram for the activated proportion of cells, $y^*/(u^*+y^*)$, with respect to the rate of CTL-mediated lysis, γ, for three different choices of spontaneous viral expression, τ. Consider an infected individual whose CTL response is characterised by a certain efficiency (thick, vertical, dotted line). For a given rate, τ, of spontaneous viral expression, the curves in the figure describe the relationship between CTL efficiency, γ, and the active proportion of proviral cells, $y^*/(u^*+y^*)$, according to model (3.1) (solid and dashed curves). Suppose that, initially, the value of τ for this individual is sufficiently low so that the active proportion lies below the theoretical threshold for pathogenesis (thin, horizontal, dotted line), thus classifying him or her as an AC (represented by the open circle ◯ in the figure). We conjecture that an increase in viral protein expression over the course of chronic HTLV-I infection could eventually push the active proportion of proviral cells over the proposed threshold for pathogenesis, leading to disease manifestation (represented by the open square □ in the figure). Here, we have selected the values of τ to be 0.03, 0.07, and 0.11 day^{-1}. All other parameters have been chosen as in Table 3.1.

ing our hypothesis on the threshold for pathogenesis is true, this suggests that either the proposed threshold for the activated proportion of proviral cells is much higher than that of most HTLV-I-infected individuals, or that perhaps τ does not change significantly over the course of chronic infection. Nevertheless, it is not clear why the rate of spontaneous viral expression τ would increase over time. Indeed, this rate is thought to be characteristic to each individual and is determined by two primary factors that remain relatively constant over time: the HTLV-I genomic sequence, which is genetically stable, and the dominance of infected T-cell clones, each defined by a specific site of integration into the host DNA (Bangham et al., 2009; Gillet et al., 2011; Meekings et al., 2008). Further studies are needed to resolve these issues and test the validity of our hypothesis.

3.10 Summary of Results from Baseline HTLV-I Model

In this chapter, we have developed a theoretical framework that incorporates sufficient biological complexity to accurately describe the fundamental within-host dynamics of HTLV-I infection and, at the same time, is able to account for some of the idiosyncrasies frequently observed in HTLV-I experimental and clinical data. Specifically, we have constructed the baseline mathematical model (3.1) to describe the *in vivo* infection dynamics of HTLV-I that incorporates two key features: (i) viral latency, and (ii) the HTLV-I-specific CTL response. The separation of infected target cells into two distinct compartments, latently infected and actively infected, is crucial to understanding the persistence of HTLV-I in the midst of a chronically stimulated anti-viral host immune response. Here we summarise the theoretical results, both mathematical and biological, that arise from investigation of model (3.1).

3.10.1 Mathematical Conclusions

To gain some insights to the behaviour of solutions to our model, we have exploited the known absolute and relative sizes of the variables to select an appropriate non-dimensional re-scaling for the model, and considered parameter values in biologically relevant ranges. Performing mathematical analysis on the non-dimensionalised system, we have identified a quantity, the basic reproduction number for viral infection R_0, as a sharp threshold parameter that determines the global dynamics of the model. Moreover, asymptotic analysis has allowed us to characterise the exponentially increasing initial transient behaviour of solutions before they level off and settle to equilibrium. The mathematical results are stated in the following two theorems in terms of the original dimensional model (3.1).

Theorem 3.4 (Global Dynamics)**.**

1) *The infection-free equilibrium P_0 always exists. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma}$ and it is globally asymptotically stable in $\bar{\Gamma}$;*

2) *if $R_0 > 1$, then the infection-free equilibrium P_0 is unstable. In addition, a unique chronic infection equilibrium $\bar{P}$ exists in $\mathring{\Gamma}$ and it is globally asymptotically stable in $\mathring{\Gamma}$.*

Theorem 3.5 (Characterisation of the Initial Transients)**.** *Suppose $R_0 > 1$ and let $\tilde{x}$, $\tilde{u}$, $\tilde{y}$, $\tilde{z}$ represent the variables in the initial phase of infection. Then, the initial transient behaviour of model* (3.1) *is characterised by*

$$\tilde{x}(t) = x_H + \big(x(0) - x_H\big)e^{-\mu_1 t}, \quad \textit{where } x_H = \frac{\lambda}{\mu_1}, \tag{3.19}$$

and

$$\begin{bmatrix} \tilde{u}(t) \\ \tilde{y}(t) \\ \tilde{z}(t) \end{bmatrix} = \vec{q_1} e^{\phi_+ t} + \vec{q_2} e^{\phi_- t} + \vec{q_3} e^{-\mu_4 t}, \tag{3.20}$$

where $\phi_\pm$ *and* $-\mu_4$ *are the eigenvalues of the linearisation of the subsystem for* u, y, z *of the dimensional model with* $x(0)$ *replacing* $\tilde{x}(t)$*, and* $\vec{q_1}$*,* $\vec{q_2}$*,* $\vec{q_3}$ *are functions of the initial conditions* $x(0)$*,* $u(0)$*,* $y(0)$*,* $z(0)$ *derived from the corresponding eigenvectors.*

Our analysis has also revealed that the timescale for the establishment of chronic infection is on the order of 10^2 days, at least in the parameter ranges under consideration. As infection by HTLV-I is typically asymptomatic, identification of chronic HTLV-I infection usually occurs long after the initial exposure to the virus, and consequently it has been difficult to study the events that occur in the initial phase of infection. Although this timescale for establishment of HTLV-I in a host has not yet been quantified, it would be interesting to see how our estimated timescale compares to that of other, more well-known, chronic (retro-) viral infections.

Establishing the global dynamics of our mathematical model (3.1) means that we have a complete understanding of the qualitative behaviour of solutions of the model. This puts us in a position where we can fully explore the model without having to worry about the occurrence of unexpected phenomena, and allows us to draw robust conclusions from further investigations of the model that rely more on computational, rather than analytical, results.

3.10.2 Biological Conclusions

From a biological perspective, we have investigated our mathematical model (3.1) in the context of several important questions regarding the persistence and pathogenesis of HTLV-I in order to gain some biologically meaningful insights into the nature of chronic HTLV-I infection. One key feature of our model that differs from previous models is the consideration of a dynamically changing within-host environment involving the complex interplay between infected target cells that may either be latent or activated, and persistent virus-specific host immune responses. Our study has shed light on the way in which HTLV-I interacts with host immunity to persist. The results of our model are consistent with experimental observations and, in addition, raise further questions that may be addressed experimentally.

In particular, we set out to illuminate the following three key questions in HTLV-I immunology.

Q1. Why is HTLV-I not silent?

▷ **Viral expression is necessary for the establishment, propagation, and persistence of HTLV-I infection.**

Recently, Asquith et al. (2007) have given a definitive resolution to the issue of HTLV-I latency by demonstrating experimentally that HTLV-I is not a fully silent infection and that on-going viral transcription is present. We asked the question: Why is HTLV-I not silent since proviral activation exposes infected target cells to immune-mediated pressure?

We investigated this question in the context of our mathematical model, and discovered that for a completely latent virus, the benefits gained by activation aided in the establishment of infection. Moreover, the rate of spontaneous expression of viral antigens was shown to be an important determinant of the proviral load.

Q2. What determines the strength of the HTLV-I-specific CTL response, and why is infection life-long?

▷ **Efficient control of the HTLV-I proviral load depends on a high rate of CTL-mediated lysis and not on the frequency of anti-HTLV-I CTLs.**

Our model has demonstrated that despite frequent viral latency, the HTLV-I-specific CTL response is a factor that determines the level of the proviral load at equilibrium. The ability of the CTL response to effectively control the proviral load may not be dependent upon the numbers of CTLs, but on how fast CTLs are able to lyse the relatively small proportion of virus-infected target cells that are activated. The results of our model agree qualitatively with experimental data from Nagai et al. (2001) and Asquith et al. (2005a).

▷ **Infection may be life-long because efficient immune responses fail to eliminate the small proportion of proviral cells that are activated.**

An efficient HTLV-I-specific CTL response kills virus-expressing actively infected cells at a rapid rate and lowers the proviral load. However, due to the benefits to the virus gained by exhibiting at least a low level of proviral expression, a plausible reason why efficient CTLs cannot eradicate the infection is because they are unable to completely eliminate the activated proportion of proviral cells that are aggressively propagating the infection. Furthermore, from our model, we have shown the somewhat non-intuitive result that a faster rate of spontaneous proviral activation and subsequent

expression of HTLV-I antigens, which exposes more infected target cells to host immunity, makes it more challenging for CTLs to achieve complete viral clearance of the actively infected target cell pool.

Q3. What characterises HTLV-I-linked pathology?

▷ **The extent of proviral activation, rather than the size of the proviral load, may distinguish clinical status and suggests a potential route for disease manifestation.**

Finding a quantifiable measure for the amount of detriment caused by HTLV-I is important in assessing the severity of infection and disease: For an asymptomatic carrier, what is the risk of developing disease? For a HAM/TSP patient, how far has the disease progressed?

A traditional measure of viral burden is the proviral load, which is a direct representation of the magnitude of the infection. However, the proviral load among both ACs and HAM/TSP patients can display large overlap and its magnitude alone cannot determine clinical status (Asquith et al., 2000; Mosley et al., 2005). We propose that the activated proportion of proviral cells, which is a direct reflection of the rate of spontaneous proviral activation and hence a measure of the "aggressiveness" of the virus itself, could be an alternative measure of viral burden that distinguishes ACs from HAM/TSP patients, independent of the size of the proviral load. Namely, we hypothesise that a high level of infected target cell activation could be associated with HAM/TSP, whereas a low level of proviral activation may correspond to asymptomatic carriage — there exists a threshold level of activation separating the two states. Our results offer insights to two unknown issues in HTLV-I pathogenesis. (i) Why do HAM/TSP treatments fail to eradicate disease? and (ii) what causes the progression of HTLV-I infection from the asymptomatic stage to HAM/TSP?

For the former, we have demonstrated using our mathematical model that treatments boosting the efficiency of the HTLV-I-specific CTL response may be inadequate at reducing the activated proportion of proviral cells below the proposed threshold level for disease, despite being effective at lowering the proviral load. For the latter, we suggest a plausible mechanism for disease manifestation: The rate of spontaneous viral activation increases over time, for example through activation induced by pro-stimulatory signalling pathways (Höllsberg, 1999), gradually raising the proportion of infected target cells that are activated and eventually crossing the proposed threshold for disease.

3.11 Modelling Development and Structural Stability

In the development of the baseline model (3.1), we understand that a range of alternative possible model structures could have been selected to represent the various host-virus interactions. Indeed, any of the terms in the model can take on a wide range of different forms, for instance, those describing the proliferation of each cell population, the transmission of virus, as well as the removal of infected cells by host immunity.

We have based the construction of the baseline model, as well as the model extensions explored in subsequent chapters, on previous within-host virus infection models, and most of the principle interactions, including production of uninfected $CD4^+$ T-cells into the system, the mechanism of infectious transmission, mitotic proviral transmission, and CTL-mediated lysis of proviral cells, are consistent with these models and are robust to slight variations in their exact functional forms (Asquith and Bangham, 2000, 2003a; Gómez-Acevedo et al., 2010; Li and Lim, 2011; Nowak and Bangham, 1996; Perelson et al., 1993; Perelson and Nelson, 1999; Wodarz et al., 1999, 2002; Wodarz, 2005). Thus, the modelling framework we have selected exhibits some degree of structural stability (though this is not explicitly quantified). For example, in the formulation of the baseline mathematical model (3.1) in Section 3.2, we have demonstrated and discussed how the exact term describing mitotic divison of the target cell population does not affect the qualitative behaviour of the model outcomes in comparison to other models (see also Appendix C.1).

A crucial difference between our approach and a number of other models is the representation of the CTL proliferation term. The expansion of CTLs is dependent on antigen stimulation and is of the form $\nu y f(z)$, where ν is the CTL proliferation rate, y is the density of proviral target cells, and $f(z)$ is a function of the CTL density, z. Common choices in within-host mathematical models of virus infection for $f(z)$ are z, $z/(z+a)$, or a more general Hill function $z^n/(z^n+a)$ for $n \geq 1$, where $a > 0$ is a positive constant. However, the precise way in which CTL proliferation should be modelled is not known. Although $CD8^+$ CTL expansion involves cellular proliferation and should therefore be proportional to the number of $CD8^+$ CTLs present, it has been noted that such proliferation should be expected to saturate; otherwise, the resulting rapid expansion of CTLs can lead to unstable dynamical behaviour (Wodarz, 2005). Following Wodarz et al. (2002) and Wodarz (2005), we consider the case $f(z) = 1$ for the functional form of $f(z)$, which is believed to be a more biologically plausible representation of CTL proliferation. As we will see later on in subsequent models, the adoption of this functional form for CTL proliferation is consistent with the co-existence of distinct epitope-specific CTL responses when we consider multiple epitopes in the following chapters.

As discussed by Asquith and Bangham (2003b), the types of theoretical results that can be obtained from a particular model are strongly influenced by the precise functional forms of the terms in the model equations. The authors point out that there is a trade-off between generality of the model and its predictive power: a more biologically realistic model aims to follow more faithfully the detailed immunological interactions and would be expected to 'correctly' capture the dynamical behaviour of the system, whereas an alternative approach would be to develop more general models that have less predictive capability but exhibit consistent behaviour independent of the choice of model. In this thesis, we lean towards the former approach and have chosen our particular modelling framework, importantly, to test the hypothesised host-virus interactions underpinning our mathematical model and, equally, to reflect what is observed in the experimental data. To address the sensitivity of the model parameters to the outputs, we have conducted a global stability analysis which has allowed us to thoroughly understand the qualitative behaviour of the mathematical model. Inevitably, a complete characterisation of the detailed immunological processes is likely unattainable, at least in the near future, due to the sheer complexity of the host-virus interaction, thereby supporting a shift towards the latter approach. Nevertheless, the modelling work presented in this thesis follows the development of a specific model structure. In future work, we would like to generalise our analyses to cover a broader and more general class of mathematical models.

3.12 Final Remarks on the Baseline Model

The baseline mathematical model (3.1) describing the host-virus interaction during persistent HTLV-I infection, which we have developed in this chapter, forms the foundation for the more complicated mathematical models in the remainder of this thesis that incorporate an increasing level of biological detail. As we have seen in our investigation of the baseline model, there exists a dynamic struggle between host immunity and infected target cells that actively display viral antigens. In fact, a variety of viral gene products are expressed by a single activated proviral cell at the same time, and due to the highly specific nature of cell-mediated immunity, each one can trigger a distinct antigen-specific CTL response. The next step is to more carefully explore this issue of simultaneous viral protein expression. In the following chapter, we examine an extension of the baseline model representing the situation where two viral proteins are simultaneously expressed, which will pave the way to further model extensions in subsequent chapters.

Chapter 4

Considering Simultaneous Expression of Two Epitopes in HTLV-I Infection: The Tax/HBZ Model

In Chapter 3 we developed and analysed a baseline mathematical model for the *in vivo* dynamics of HTLV-I infection in the presence of persistent host immunity (Lim and Maini, 2014). Mathematical and numerical investigations of this baseline HTLV-I model, given by the system of equations (3.1), yielded predictions that help to resolve four main gaps in the understanding of HTLV-I persistence. First, HTLV-I is not a transcriptionally silent infection because viral expression is required for establishing the infection and determining the equilibrium proviral load. Second, the abundance of anti-HTLV-I CTLs is maintained by a rapid rate of CTL expansion in response to a high, continuous level of viral antigen expression. Moreover, a high CTL frequency is not responsible for controlling the proviral load; rather, CTL quality depends on how fast host immunity eliminates activated proviral cells. Third, it is possible that the reason why infection is so difficult to clear is not because most of the virus is latent, but because CTLs cannot eradicate the small proportion of proviral cells that are activated. Fourth, we can reconcile the experimental observation that the magnitude of an infected individual's proviral load is by itself not able to classify clinical status (AC or HAM/TSP patient) by suggesting that the extent of proviral activation may be an important measure of viral detriment. Using this alternative measure of viral detriment, we suggested a promising way not only to distinguish the clinical status of chronically infected individuals independent of the magnitude of the proviral load, but also propose a plausible quantifiable mechanism for the progression of the inflammatory disease, HAM/TSP. The baseline mathematical model (3.1) has therefore helped to improve our understanding as to why HTLV-I has evolved mechanisms to retain transcriptional

activity rather than adopting a life-cycle that is completely latent. Indeed, it is becoming increasingly apparent that the relatively small fraction of infected target cells that actively express viral antigens in a sea of proviral cells that is largely latent has a strong effect on HTLV-I persistence.

Yet, there still remain many unresolved issues in current understanding of HTLV-I to which our model cannot provide insights. Indeed, in light of recent experimental work, it is evident that the dynamics of HTLV-I infection exhibit a much more sophisticated structure than our baseline model (3.1) has described. To try and address these more detailed issues concerning HTLV-I infection and persistence, we must therefore consider a more biologically detailed manifestation of our mathematical model.

4.1 Motivation

At this point, we are reminded of our central theme of understanding viral persistence in the presence of host immunity. Infected target cell activation is beneficial to the virus, but how is it able to make any meaningful contribution to the long-term persistence of the infection when HTLV-I-specific CTLs rapidly eliminate infected cells that express viral proteins? Do actively infected cells simply wait to be removed by host immunity? Since the adaptive immune response tends to improve over the course of chronic viral infection (Parham, 2005), and HTLV-I infection does not significantly impair immune functionality (Asquith and Bangham, 2007), it would be expected that strong CTL pressure would eventually drive the infection in the direction of latency. This is clearly not what happens, as a high level of continuous expression can certainly accompany a large, chronically activated CTL response (Asquith et al., 2000; Bangham, 2003). Moreover, insights from our baseline model (3.1) have supported a positive role for viral activation in the persistence of HTLV-I.

The question that arises from the above observations then is: How do HTLV-I-infected cells evade CTL-mediated elimination whilst retaining transcriptional activity? A plausible explanation comes to light when we realise that the expression of multiple viral antigens means that the same pool of actively infected cells can stimulate multiple distinct anti-HTLV-I immune responses. It is reasonable to expect a varying array of stimulation for each of the different antigen-specific immune responses, resulting in a heterogeneous CTL population structure.

In this chapter, we extend the baseline model (3.1) to incorporate a biologically important aspect of HTLV-I infection, namely that infected target cell activation is accompanied by the simultaneous expression of multiple HTLV-I proteins. Despite there being a variety of distinct HTLV-I proteins, roughly 12 in total (Boxus and Willems, 2009; MacNamara et al., 2010), to begin our investigation regarding the presence of multiple viral protein targets, we

first focus on the expression of only two epitope targets here. Although our model does not distinguish between the various HTLV-I proteins in terms of their respective functions, to all intents and purposes in the two-epitope case we will call the viral proteins Tax and HBZ since these are believed to be the two proteins that play the greatest role in determining the course of chronic HTLV-I infection (Boxus and Willems, 2009; MacNamara et al., 2010). Consequently, we refer to the two-epitope model developed in this chapter as the Tax/HBZ model. In the following chapter, we generalise our results to cover any arbitrary number of CTL epitopes, which gives rise to what we call the (general) multiple epitope model for HTLV-I infection.

We use the Tax/HBZ model, which is given by the system of equations (4.1) to address the following issues that could not have been resolved using previous models:

(1) Which factors contribute to the prevalence of a certain CTL type when all target antigens can be recognised by host immunity?

(2) Why can one CTL response have a greater impact on the proviral load than another? More broadly, what defines a 'good' epitope-specific immune response?

4.2 Extension of Baseline Model to a Basic Multiple Epitope Model: The Tax/HBZ Model

Motivated by the above discussions, we formulate and analyse the Tax/HBZ model (4.1), which is a direct extension of our baseline model (3.1) that distinguishes between different anti-HTLV-I CTL responses having specificities to only one of the two simultaneously expressed viral antigens, Tax or HBZ.

4.2.1 Tax/HBZ Model Formulation

We separate the T-cell populations into two broad classes of target cells and immune cells. Target cells are $CD4^+$ helper T-cells of which there are three types, which we identify according to their properties: healthy, latently infected, and actively infected. Meanwhile, the immune cell pool of $CD8^+$ CTLs consists of two types depending on antigen specificity: Tax-specific and HBZ-specific. Denote

$x(t)$: number of healthy (uninfected) $CD4^+$ helper T-cells at time t,

$u(t)$: number of latently infected $CD4^+$ helper T-cells at time t,

$y(t)$: number of actively infected $CD4^+$ helper T-cells at time t,

$z(t)$: number of Tax-specific $CD8^+$ CTLs at time t,

$s(t)$: number of HBZ-specific $CD8^+$ CTLs at time t.

The incorporation of the simultaneous expression of Tax and HBZ on activated proviral cells necessitates the consideration of multiple CTL responses that are specific to each particular epitope or antigenic target. It is known that actively infected T-cells displaying multiple viral epitopes concurrently can be recognised by more than one CTL response at a time (Macken and Perelson, 1984; Zagury et al., 1979). Upon proviral activation (which is accompanied by the expression of all viral antigens), antigen-specific CTL responses, either Tax-specific or HBZ-specific or both, engage the infected T-cell and the corresponding CTL type undergoes two processes: (i) it is stimulated to proliferate, and (ii) it induces cell death in the proviral cell to which it is engaged. The effect of proviral activation on the interactions between infected T-cells and antigen-specific CTLs, both Tax- and HBZ-specific, are illustrated for reference in Figure 4.1.

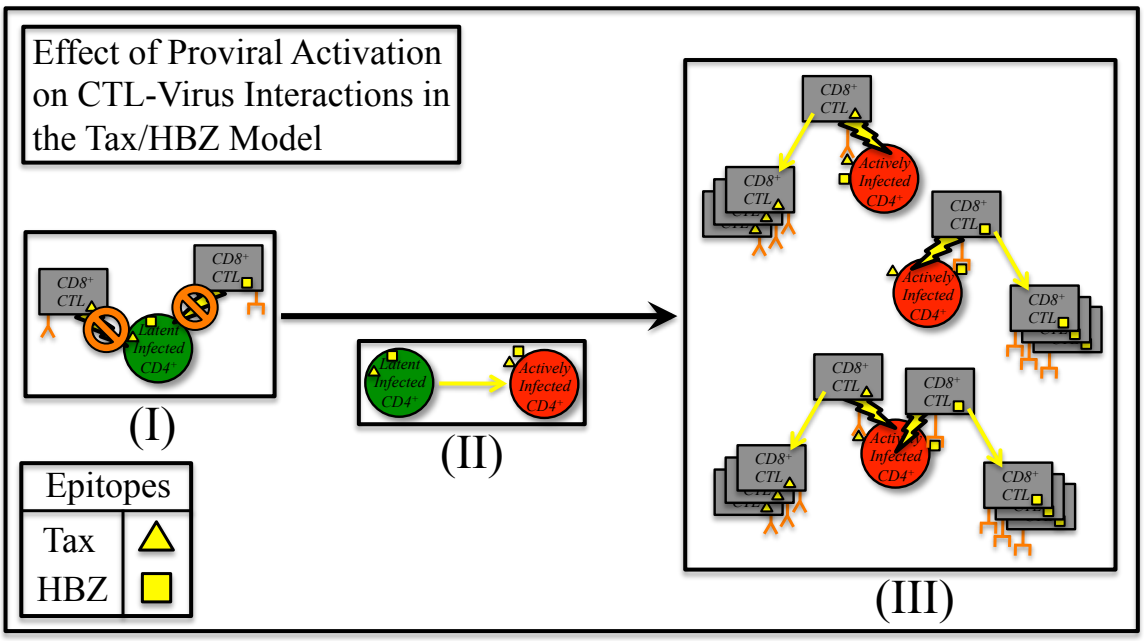


Figure 4.1: The effect of proviral activation on CTL-virus interactions in the Tax/HBZ model (4.1). **(I)** Proviral cells that are transcriptionally latent are able to hide from immune surveillance and avoid recognition and subsequent destruction by the host immune response. **(II)** Proviral activation of infected target cells is accompanied by the simultaneous expression of both Tax and HBZ. **(III)** Host cellular immunity is able to recognise each of the two viral proteins. Any given actively infected target cell can stimulate the expansion of either antigen-specific CTL response. The Tax and HBZ epitopes are illustrated in the figure by the yellow shapes as indicated.

To develop the Tax/HBZ model, we combine the T-cell dynamics from the baseline model with the more detailed antigen-specific CTL-virus interactions described above. An illustration of the various interactions that take place among the target and immune cell sub-populations in our two epitope Tax/HBZ model is shown in Figure 4.2.

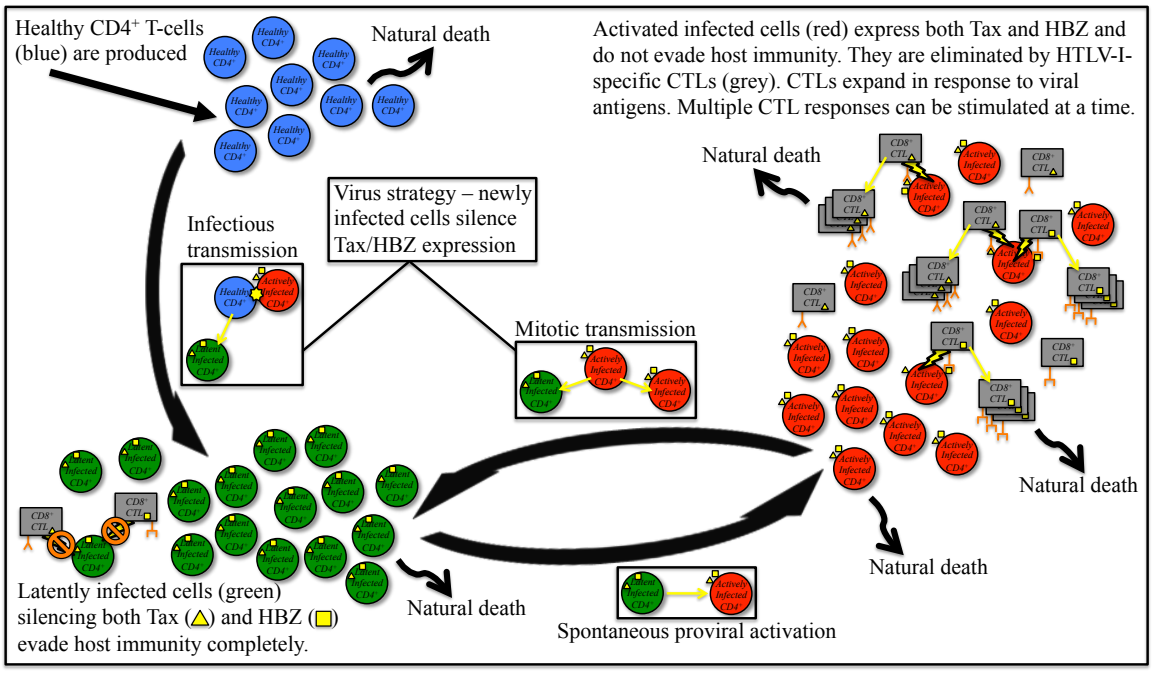


Figure 4.2: A schematic representation of the simultaneous expression of the viral proteins Tax and HBZ that motivates the formulation of a more detailed mathematical model for HTLV-I infection *in vivo*. The corresponding dynamical system is represented by the non-linear set of ordinary differential equations (4.1).

The dynamics of the Tax/HBZ model is represented by the following non-linear system of ordinary differential equations (4.1).

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \overbrace{\lambda}^{\text{T-cell production}} - \overbrace{\beta x y}^{\text{infectious transmission}} - \overbrace{\mu_1 x}^{\text{natural death}} \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= \overbrace{\beta x y}^{\text{infectious transmission}} + \overbrace{r y}^{\text{mitotic transmission}} - \overbrace{(\tau + \mu_2) u}^{\text{spontaneous activation and natural death}} \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \overbrace{\tau u}^{\text{spontaneous activation}} - \overbrace{\gamma y z}^{\text{Tax-specific CTL lysis}} - \overbrace{\gamma' y s}^{\text{HBZ-specific CTL lysis}} - \overbrace{\mu_3 y}^{\text{natural death}} \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= \overbrace{\delta \nu y}^{\text{CTL proliferation}} - \overbrace{\mu_4 z}^{\text{natural death}} \\
\frac{\mathrm{d}s}{\mathrm{d}t} &= \overbrace{(1-\delta)\nu' y}^{\text{CTL proliferation}} - \overbrace{\mu_5 s.}^{\text{natural death}}
\end{aligned}
\tag{4.1}
$$

Many of the underlying processes of the infection dynamics follow the formulation of the baseline HTLV-I model (3.1) (refer to Chapter 3 as well as to Lim and Maini, 2014). However, what is different here is the refinement of the anti-HTLV-I CTL response to reflect specificities to each of the two HTLV-I proteins, Tax or HBZ. These mechanisms will be described more thoroughly.

Infectious transmission of the provirus occurs as a result of direct cell-to-cell contact

between a healthy cell $x(t)$ and an actively infected cell $y(t)$ with coefficient of infectious transmissibility β. Actively infected cells, $y(t)$, undergo rapid, mitotic division proportional to their numbers and their expansion is assumed to obey an exponential growth profile with proliferation rate r. Newly infected target cells from either route of transmission are assumed to immediately silence viral expression, thus entering the latently infected target cell pool, $u(t)$. A small fraction of transcriptionally silent target cells spontaneously become re-activated at a rate τ, simultaneously expressing both Tax and HBZ in the $y(t)$ compartment. These actively infected cells both stimulate the proliferation of, and are eliminated by, the two subpopulations of anti-HTLV-I $CD8^+$ CTLs, those that target Tax and those that target HBZ, and CTL-mediated lysis is assumed to have a bilinear incidence form. Tax-specific CTLs eliminate such cells at a rate γ, while the rate of lysis due to HBZ-specific CTLs is γ'.

Since actively infected cells $y(t)$ express both viral antigens, they stimulate the expansion of both populations of anti-HTLV-I CTLs. We assume that the proliferation rate of Tax-specific CTLs $z(t)$ is linear with respect to the abundance of activated proviral cells at a rate constant ν, and similarly the proliferation rate of HBZ-specific CTLs $s(t)$ is ν'. Although activated proviral cells can simultaneously initiate the expansion of both types of CTLs, a proportion of all infected T-cell and CTL interactions engage a particular CTL type and stimulate its proliferation. We represent this aspect of the host-virus interaction by a parameter δ, where $0 \leq \delta \leq 1$, so that the fraction $\delta y(t)$ of fully activated proviral cells stimulate Tax-specific CTL growth, whereas the remaining fraction $(1-\delta)y(t)$ stimulate HBZ-specific CTL growth. In fact, the parameter δ captures an important concept in immunology known as immunogenicity, which is the ability of a substance to provoke an adaptive immune response (Parham, 2005). Following Nowak (1996), let us define the immunogenicity of a CTL epitope to be the rate at which this epitope induces CTL proliferation. Then, in the context of model (4.1), the immunological meaning of the parameter δ is precisely that of immunogenicity of the Tax antigen. Meanwhile, since δ is a proportion lying in between 0 and 1, then its complement $(1-\delta)$ represents the immunogenicity of the HBZ antigen. Hence, large values of δ (i.e. close to 1) corresponding to a strong stimulation of Tax-specific CTLs and a weak stimulation of HBZ-specific CTLs is equivalent to high Tax immunogenicity and low HBZ immunogenicity. The reverse is true for small values of δ (i.e. close to 0), namely, low Tax immunogenicity and high HBZ immunogenicity. This has significant biological implications which will be explored in more detail later.

Lastly, healthy, latently infected, and actively infected target cells are removed from the system at linear rates μ_1, μ_2 and μ_3, respectively, whereas Tax-specific and HBZ-specific

CTLs have respective removal rates μ_4 and μ_5. All parameters are assumed to be positive, with the exception of δ, which lies in the closed interval between 0 and 1.

Remark 4.1. The Tax/HBZ model is a natural, direct extension of the baseline model. In the situation where there is no distinction between the two types of anti-HTLV-I CTLs so that $\gamma = \gamma'$, $\nu = \nu'$, and $\mu_4 = \mu_5$, then letting $\tilde{z}(t)$ denote the total number of anti-HTLV-I CTLs $z(t) + s(t)$, we observe that the model reduces to the baseline model (3.1) analysed in Chapter 3 (see also Lim and Maini, 2014).

Remark 4.2. In the formulation of model (5.1), we incorporate the role of epitope immunogenicity through the parameter δ representing the proportion of infected target cells and CTL interactions that engage Tax-specific CTLs. Although there is a possibility that multiple CTLs can engage a given actively infected target cell (displaying both viral epitopes) at the same time, this will be reflected in the parameter δ.

4.2.2 Virus Fitness and Heterogeneity in CTL Responses

The basic reproduction number R_0 with respect to model (4.1) is

$$R_0 = \frac{\tau}{\mu_3(\tau + \mu_2)}\Big(\beta x_H + r\Big), \quad \text{where } x_H = \frac{\lambda}{\mu_1}. \tag{4.2}$$

Note that this is precisely the same expression for the basic reproduction number R_0 for the baseline HTLV-I model (3.1). Meanwhile, there can also exist heterogeneity in epitope-specific CTL responses, which depend on factors relating to effector functionality as well as growth and death properties of each CTL type. For the former, effector function is defined by the rates of CTL-mediated lysis (γ, γ'). For the latter, each CTL type is governed by intrinsic properties such as its rate of proliferation (ν, ν') and lifespan ($1/\mu_4$, $1/\mu_5$). We group together the growth and death rates into a single quantity ϕ representing what we will refer to as the *effective CTL expansion rate* for each epitope-specific CTL response:

$$\phi_{\text{epitope}} = \frac{\nu_{\text{epitope}}}{\mu_{\text{epitope}}}. \qquad \Big[\text{Effective CTL Expansion Rate}\Big] \tag{4.3}$$

With respect to the Tax/HBZ model (4.1), the effective CTL expansion rates for Tax-specific and HBZ-specific CTL types are, respectively,

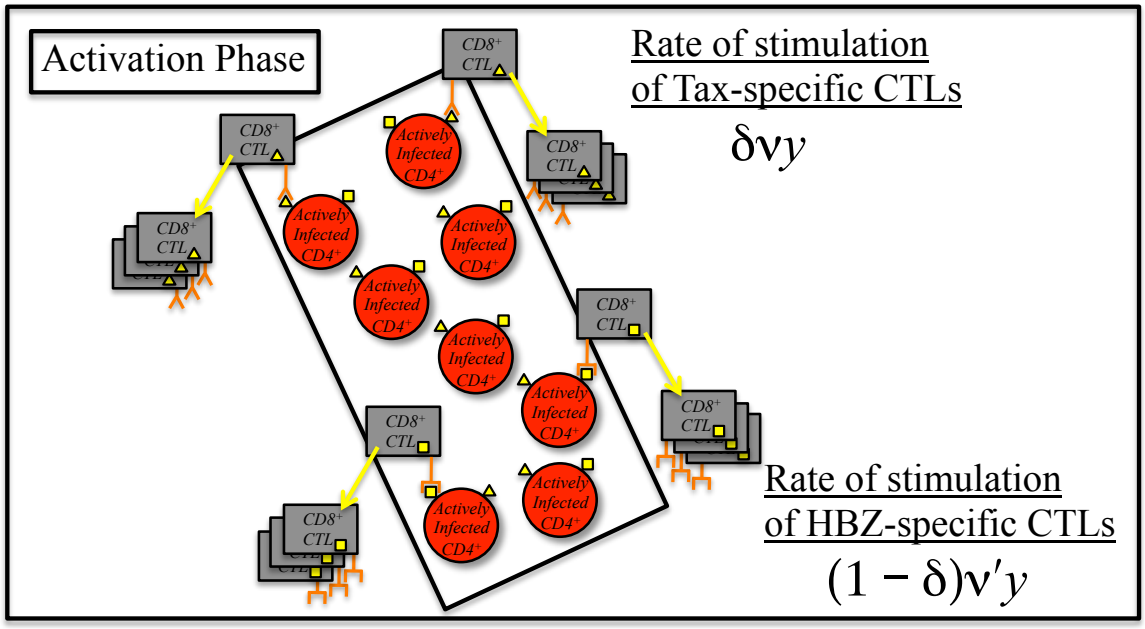
$$\phi_{\text{Tax}} = \frac{\nu}{\mu_4} \qquad \text{and} \qquad \phi_{\text{HBZ}} = \frac{\nu'}{\mu_5}. \tag{4.4}$$

4.2.3 A Detailed Look at the Activation and Effector Phases of the Cell-Mediated Immune Response

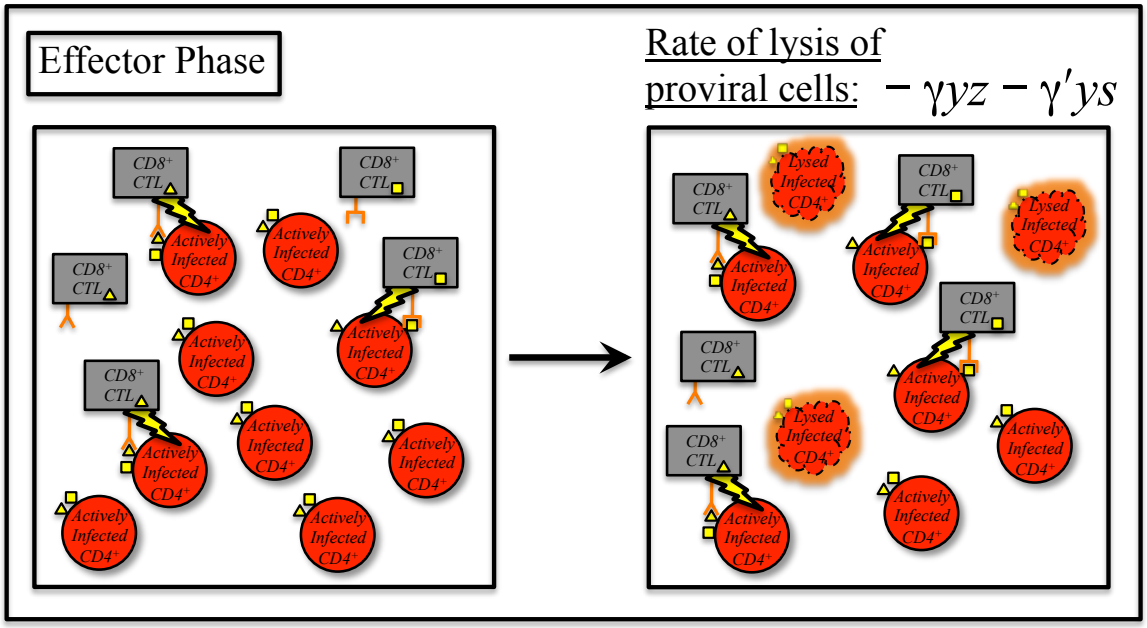
In order to get a sense of the dynamic that our Tax-HBZ model (4.1) is meant to capture, we have to delve a bit deeper into one of the details of cell-mediated immunity, whilst still maintaining as much simplicity in the underlying mechanisms as possible. In particular, we acknowledge that there are two distinct phases involved in the production of an antigen-specific cellular immune response: (i) the activation phase, and (ii) the effector phase. Because CTLs interact with the pool of proviral cells in a different way during the effector phase than during the activation phase, in our modelling approach, we conceptually draw a distinction between these two phases of cell-mediated immunity.

During the activation phase, fragments of viral proteins are presented by professional antigen-presenting cells (APCs) such as dendritic cells, macrophages, and B-cells to naïve $CD8^+$ T-cells in the secondary lymphoid tissue (usually in the lymph nodes) (Parham, 2005). T-cells that are able to recognise the foreign protein then differentiate into activated CTLs which are specific for that particular antigen and subsequently undergo clonal expansion. In terms of the Tax/HBZ model (4.1), the activation phase corresponds to the recognition of viral antigens by CTLs, thereby stimulating their proliferation at a rate directly proportional to the concentration of antigen (i.e. CTL proliferation grows exponentially with respect to the numbers of proviral cells present). For this phase, we assume that all infected target cells are recognised by at least one of the two CTL types: the proportion $\delta y(t)$ of the proviral population stimulates the expansion of Tax-specific CTLs at the rate ν, whereas the remaining proportion $(1-\delta)y(t)$ stimulate HBZ-specific CTL proliferation at the rate ν'. The expanded CTL populations then proceed to the effector phase of cellular immunity. A schematic diagram of the activation phase is illustrated in Figure 4.3(a).

The effector phase describes the stage where cell-mediated lysis of proviral cells takes place. Differentiated and now activated epitope-specific CTLs, known also as effector T-cells, bind and eliminate infected target cells expressing the appropriate cognate antigen by inducing apoptosis (i.e. programmed cell death) via one of two possible pathways. The first pathway is the directed release of cytotoxins such as perforin, granzymes and granulysin by the effector CTL onto the surface of the target cell, causing fatal damage to the recipient proviral cell. The second pathway involves the interaction of receptors on the effector and target cells, namely the cell-surface cytokine protein Fas ligand on the CTL and the Fas molecules on the surface of the target cell, which leads to apoptosis of the infected cell (Parham, 2005). Regardless of the pathway chosen, after apoptosis has been induced in the target cell, the effector T-cell is able to move on to the next infected cell, meaning that a single CTL can lyse several target cells in succession (Parham, 2005). In terms of



(a) CTL activation phase.



(b) CTL effector phase.

Figure 4.3: A more detailed look at the two phases of cell-mediated adaptive immunity: the activation phase and the effector phase. (a) During the activation phase, CTLs recognise and bind to their respective cognate antigens and undergo clonal expansion. With respect to the Tax/HBZ model (4.1), we suppose that a fraction $\delta y(t)$ of the proviral population stimulates the proliferation of Tax-specific CTLs at a rate ν, whereas the remaining fraction $(1-\delta)y(t)$ stimulates the growth of HBZ-specific CTLs at a rate ν'. (b) During the effector phase, antigen-specific CTLs which have become activated (known as effector CTLs) eliminate actively infected target cells by inducing apoptosis (or programmed cell death). In the Tax/HBZ model, we assume that such cell-mediated lysis occurs according to mass action kinetics by Tax-specific and HBZ-specific CTLs at respective rates γ and γ'.

the Tax/HBZ model (4.1), in the effector phase, we make the assumption that the entire population $y(t)$ of virus-expressing infected target cells is subject to immune-mediated lysis by either of the two types of epitope-specific CTL in line with the law of mass action. Thus, we describe the rate of immune-mediated lysis of activated proviral cells according to the term $-\gamma yz - \gamma' ys$. A schematic illustration of the effector phase is shown in Figure 4.3(b).

4.2.4 Global Dynamical Behaviour of the Tax/HBZ Model

We observe from the model equations that for any set of non-negative initial conditions, $x(t), u(t), y(t), z(t), s(t) \geq 0$ for all $t \geq 0$. Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$, $\tilde{\nu} = \max\{\nu, \nu'\}$, and $\tilde{\mu} = \min\{\mu_4, \mu_5\}$, and consider the closed, bounded region

$$\Gamma := \left\{ (x, u, y, z, s) \in \mathbb{R}^5_+ : x \leq \frac{\lambda}{\mu_1},\ x + u + y \leq \frac{\lambda}{\mu - r},\ z + s \leq \frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu - r)} \right\}.$$

Theorem 4.1. *Assume that $r < \mu$. Then, the set Γ defines a positively invariant region with respect to model* (4.1)*. All solutions are bounded for $t \geq 0$ and eventually enter Γ.*

The following theorem demonstrates that the basic reproduction number R_0 is a sharp threshold parameter that establishes the global dynamical behaviour of model (4.1).

Theorem 4.2 (Global Dynamics of the Tax/HBZ Model)**.**

1) *The infection-free equilibrium P_0 always exists. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma}$ and it is globally asymptotically stable in $\bar{\Gamma}$;*

2) *if $R_0 > 1$, then the infection-free equilibrium P_0 is unstable. In addition, a unique chronic infection equilibrium P^* exists in $\mathring{\Gamma}$ and it is globally asymptotically stable in $\mathring{\Gamma}$.*

The proofs of Thereom 4.1 and Theorem 4.2 follow a similar approach as for the baseline model (3.1) and have therefore been omitted. The principal differences in the proofs arise due to the consideration of varying epitope immunogenicities resulting from immune selection targeting either Tax or HBZ. We will generalise the two-epitope model (4.1) to the situation with an arbitrary number, $p \in \mathbb{N}$, of epitopes in Chapter 5. The above analytical results thus correspond to the special case when $p = 2$ in Theorems 5.1 and 5.2.

4.3 Differences in the Effective Expansion Rate and Functionality Between Epitope-Specific CTL Responses Determine How Immunogenicity Affects Measures of CTL and Virus Abundance

A key feature of the Tax/HBZ model (4.1) which distinguishes it from the baseline model (3.1) is that actively infected target cells express both Tax and HBZ proteins simultaneously and, furthermore, host cellular immunity has the ability to differentiate between the two viral antigens, hence implying the existence of two distinct types of immune responses, Tax-specific and HBZ-specific. The greatest advantage to this extended approach is that we can investigate how diversity in the CTL population can impact various aspects of the host-virus interaction.

4.3.1 CTL Dominance and CTL-to-Virus Ratio are Dependent on Immunogenicity and Factors Related to CTL Proliferation and Lifespan, but not Efficiency

There are two traditional metrics to quantify the magnitude of the cellular immune response: (i) total CTL abundance, which we will also refer to as *CTL dominance*, and (ii) CTL abundance relative to the size of the virus compartment. Because multiple CTL types are present, we adjust the aforementioned definitions in order to distinguish between the two immune responses. For the former, to keep track of the two different CTL pools, we consider the dominance of a particular CTL, that is, the proportion or relative abundance of each CTL type. For the latter, we consider the CTL to virus ratio for a particular antigen-specific immune response, or in other words, the ratio of each CTL type with respect to the total numbers of virus-expressing target cells.

We explore in more detail how the two CTL 'fitness' factors, namely immunogenicity and the effective CTL expansion rate of Tax-specific and HBZ-specific CTLs, contribute to each of the two CTL measures. We discover that both a high prevalence of CTLs and a large CTL to virus ratio for a particular immune response are the result of at least a high relative effective expansion rate or a high degree of epitope immunogenicity.

CTL Dominance

The dominance of epitope-specific CTLs does not depend on its efficiency (i.e. rate of CTL-mediated lysis), but only on immunogenicity, CTL proliferation and CTL lifespan. Specifically, the relative abundances of Tax- and HBZ-specific CTLs are given by the proportions of each CTL type's effective expansion rate scaled by its respective immunogenicity (see

Appendix A.2.1 for a full derivation of the expression for dominance of Tax-specific CTLs). That is,

$$\frac{z^*}{z^*+s^*} = \frac{\delta\phi_{\text{Tax}}}{\delta\phi_{\text{Tax}} + (1-\delta)\phi_{\text{HBZ}}} \quad \text{and} \quad \frac{s^*}{z^*+s^*} = \frac{(1-\delta)\phi_{\text{HBZ}}}{\delta\phi_{\text{Tax}} + (1-\delta)\phi_{\text{HBZ}}}, \tag{4.5}$$

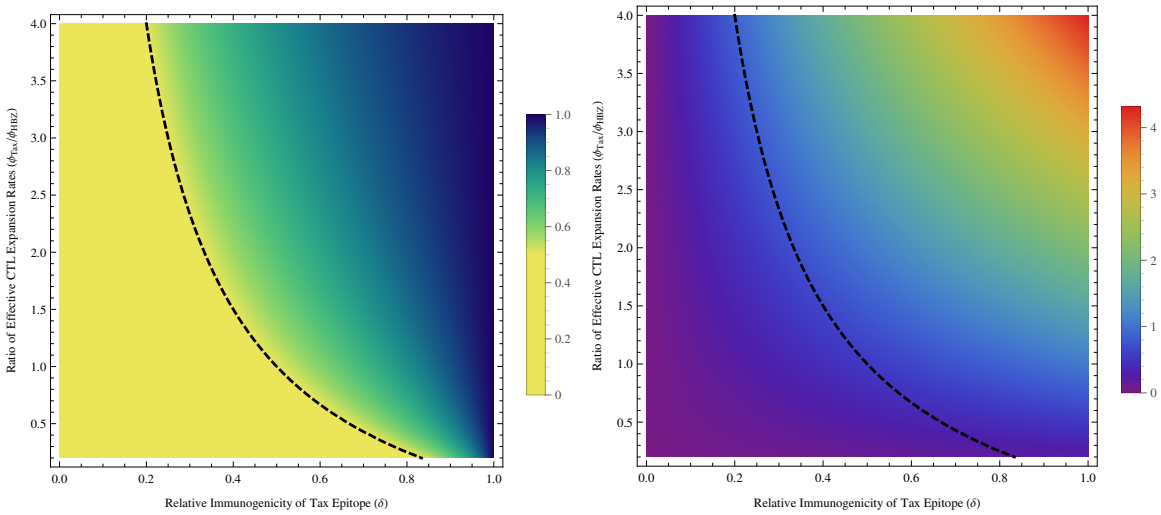
$$\Big[\text{CTL Dominance}\Big]$$

where $\phi_{\text{Tax}} = \nu/\mu_4$ and $\phi_{\text{HBZ}} = \nu'/\mu_5$ are the growth/death ratios, or effective CTL expansion rates, of each of the two types of CTLs as defined in Equation (4.3). We point out that in accordance with Equations (4.5), the strength of dominance of Tax-specific CTLs is the consequence of higher relative immunogenicity towards the Tax epitope (i.e. larger δ) as well as a higher effective CTL expansion rate of Tax-specific CTLs over CTLs that target HBZ (i.e. higher ϕ_{Tax} compared to ϕ_{HBZ}). In the extreme cases when there is complete immunogenicity towards one of the epitopes, in particular when either $\delta = 1$, corresponding to Tax being solely immunogenic, or $\delta = 0$, corresponding to HBZ being solely immunogenic, the cellular immune effector pool consists only of one CTL type. In between, the two types of CTL responses, Tax-specific and HBZ-specific, co-exist at various prevalences and achieve equal equilibrium abundances when $\delta = \tilde{\delta}$, where $\tilde{\delta}$ is defined as follows:

$$\tilde{\delta} := \frac{1}{\phi_{\text{Tax}}/\phi_{\text{HBZ}} + 1}. \tag{4.6}$$

Since the prevalence of each CTL type is determined solely by its relative immunogenicity (which here is δ for Tax and $(1-\delta)$ for HBZ) and its effective rate of CTL expansion (ϕ_{epitope}), we can think of these two factors as defining the 'fitness' of each CTL response. Figure 4.4(a) illustrates a heat map of Tax-specific CTL dominance and how it is influenced by the relationship between the two CTL 'fitness' factors. Along the horizontal axis is the relative immunogenicity (δ) of the Tax epitope. Note that its complement, $(1-\delta)$, describes the relative immunogenicity of the HBZ epitope. Meanwhile, the vertical axis indicates the ratio of effective CTL expansion rates between the two types of CTL responses. The dashed line is precisely the curve $\delta = \tilde{\delta}$ from Equation (4.6) along which the two CTL types exist at equal abundances. The strength of dominance, in terms of relative abundance within the CTL pool, of Tax-specific CTLs is indicated by a colour gradient lying to the right of the dashed line (i.e. $\delta > \tilde{\delta}$).

We make the following observations directly from Equation (4.6). First, when there is a skew towards Tax-specific CTLs so that $\phi_{\text{Tax}}/\phi_{\text{HBZ}} > 1$, then $\tilde{\delta} < 1/2$, meaning that lower Tax immunogenicity is required for Tax-specific CTLs to dominate the CTL pool. Conversely, a skew towards HBZ-specific CTLs such that $\phi_{\text{Tax}}/\phi_{\text{HBZ}} < 1$ implies that $\tilde{\delta} > 1/2$, which means that a higher level of Tax immunogenicity is required for Tax-specific



(a) Tax-specific CTL dominance.

(b) Tax-specific CTL-to-virus ratio.

Figure 4.4: Heat maps showing (a) the strength of CTL dominance and (b) the CTL to active virus ratio of Tax-specific CTLs in the two-epitope Tax/HBZ model (4.1). CTL dominance of each CTL type, Tax-specific or HBZ-specific, can be described analytically by the expressions in Equation (4.5). Meanwhile, the CTL-to-virus ratios (with respect to either Tax- or HBZ-specific immunity) are given by the expressions in Equation (4.7). Both of these measures of CTL prevalences are determined by only two factors: (i) the ratio between the effective CTL expansion rates for each immune response $\left(\phi_{\text{Tax}}/\phi_{\text{HBZ}}\right)$, and (ii) the relative immunogenicity of each epitope target (δ). In the heat maps above, we have plotted the two CTL measures with respect to the pool of Tax-specific CTLs; however, the corresponding heat maps with respect to the subset of HBZ-specific CTLs are analogous to the figures shown here. The dashed line represents the curve with respect to these two factors along which Tax-specific and HBZ-specific CTLs are equal in terms of either of the two measures of CTL abundance as described by Equation (4.6). Parameter values used to generate the above plots lie within physiologically realistic ranges as discussed in Table 3.1, with the exception of the parameter δ representing the immunogenicity of the Tax epitope, which is a fraction that can vary between 0 and 1.

CTLs to exhibit dominance. In the neutral case where both CTL responses have the same effective expansion rate, i.e. $\phi_{\text{Tax}}/\phi_{\text{HBZ}} = 1$, the boundary between Tax-specific and HBZ-specific CTL dominance occurs when the immunogenicities of each epitope are equal, i.e. $\tilde{\delta} = 1/2$.

CTL to Virus Ratio

From an experimental viewpoint, there appears to be no definitively established relationship between the abundance of CTLs and proviral load. Indeed, it has been observed in sera from different HTLV-I sero-positive individuals that the frequency of CTLs and the magnitude of the proviral load can display either positive, negative, or no correlation between the two

measures (Asquith et al., 2005a; Goon et al., 2004a; Kubota et al., 2000; Nagai et al., 2001; Rowan and Bangham, 2011; Wodarz et al., 2001). In our baseline model (3.1), as well as in the Tax/HBZ model (4.1), the production of virus-specific CTLs is directly related to the size of the pool of activated proviral target cells, suggesting that a high abundance of CTLs can be indicative of a high, rather than low, proviral load, which is in qualitative agreement with experimental data from Nagai et al. (2001), as we have demonstrated previously in Chapter 3 using the results from our baseline model (3.1). However, it is instructive to think of a measure of the CTL abundance scaled with respect to the size of the proviral load. We therefore consider the ratio of CTLs to active virus to gauge how many effector CTLs are present per proviral cell. The CTL to active virus ratio for Tax-specific and HBZ-specific CTL responses, respectively, are bi-linear functions of the ratio of effective expansion rates $\left(\phi_{\text{Tax}}/\phi_{\text{HBZ}}\right)$ and the relative immunogenicity (δ), and from the latter two equations in model (4.1) are given by:

$$\frac{z^*}{y^*} = \delta\phi_{\text{Tax}} \qquad \text{and} \qquad \frac{s^*}{y^*} = (1-\delta)\phi_{HBZ}. \qquad \left[\text{CTL-To-Virus Ratio}\right] \tag{4.7}$$

A heat map illustrating the CTL to virus ratio for Tax-specific CTLs with respect to the extent of Tax immunogenicity (δ) and the ratio of the effective expansion rates between the two types of cellular immune responses $\left(\phi_{\text{Tax}}/\phi_{\text{HBZ}}\right)$ is shown in Figure 4.4(b). From both Equations (4.7) and Figure 4.4(b), we see that a high CTL to virus ratio for a given CTL type (for example, Tax-specific CTLs) is the result of a high degree of immunogenicity coupled with a rapid rate of effective CTL expansion. The dashed line traces out the curve along which the CTL-to-virus ratios are equal for both Tax-specific and HBZ-specific CTL responses, and is equivalent to the curve representing equal abundances for the two CTL types as in Equation (4.6).

The ratio of all CTLs to active virus is a sum of the above two quantities, and yields the following expression.

$$\frac{z^* + s^*}{y^*} = \delta\phi_{\text{Tax}} + (1-\delta)\phi_{HBZ}. \qquad \left[\text{Total CTL-To-Virus Ratio}\right] \tag{4.8}$$

From Equation (4.8), it is clear that the CTL to active virus ratio is an increasing function of both ϕ_{Tax} and ϕ_{HBZ}.

4.3.2 The Relationship Between Proviral Load and CTL-Related Factors, Namely the Effective CTL Ratio and Immunogenicity

Differences between the two types of anti-HTLV-I CTL responses, not only in their respective effective CTL expansion rates (i.e. growth and death), but also in their rates of lysis,

can have a differential impact on the direction in which immunogenicity influences the magnitude of the equilibrium proviral load. We consider the following quantity, η, which defines what we call the *effective ratio* between the two types of CTL responses, one targeting Tax and the other targeting HBZ:

$$\eta = \frac{\gamma}{\gamma'}\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}. \qquad \left[\text{Effective Ratio}\right] \tag{4.9}$$

The immunogenicity towards the Tax epitope could either contribute to or restrict the level of virus depending on the value of the effective ratio η. The following theorem makes this idea more rigorous in terms of our mathematical model (4.1). For the proof, we refer the reader to Appendix A.2.2.

Theorem 4.3. *Assume that $R_0 > 1$. Then, there are three possible scenarios in which the immunogenicity δ of the Tax epitope can impact viral abundance in terms of the equilibrium proviral load (PVL), v^*, depending on the effective ratio η between the two types of epitope-specific CTL responses.*

1. *If $\eta < 1$, then $\dfrac{\mathrm{d}v^*}{\mathrm{d}\delta} > 0$, i.e. PVL displays positive correlation with Tax immunogenicity;*
2. *If $\eta = 1$, then $\dfrac{\mathrm{d}v^*}{\mathrm{d}\delta} = 0$, i.e. PVL displays no correlation with Tax immunogenicity;*
3. *If $\eta > 1$, then $\dfrac{\mathrm{d}v^*}{\mathrm{d}\delta} < 0$, i.e. PVL displays negative correlation with Tax immunogenicity.*

Figure 4.5 visually illustrates the theoretical result from Theorem 4.3 by showing a heat map for the PVL (v^*) with respect to the immunogenicity of Tax (δ) along the horizontal axis and the effective ratio (η) between the two CTLs along the vertical axis. We briefly highlight that, unlike the relative abundance of CTLs and the CTL to virus ratio, the PVL is dependent on the efficiency of CTL-mediated lysis (i.e. γ, γ') in addition to its dependence on the effective expansion rates for each CTL type (i.e. ϕ_{Tax}, ϕ_{HBZ}).

On the one hand, by taking horizontal cross-sections of the heat map in Figure 4.5, we observe the outcome of the theoretical result from Theorem 4.3 above. Namely, that depending on whether the effective ratio between the two CTL responses is less than, equal to, or greater than unity, the PVL v^* would, respectively, be an increasing, neutral, or decreasing function of the extent of immunogenicity δ of the Tax epitope. Indeed, we observe that a skew in the effective ratio towards anti-Tax CTLs (i.e. $\eta < 1$) results in a positive relationship between Tax immunogenicity and the size of the PVL. Meanwhile, the opposite is true if the effective ratio is skewed towards anti-HBZ CTLs (i.e. $\eta > 1$). At the

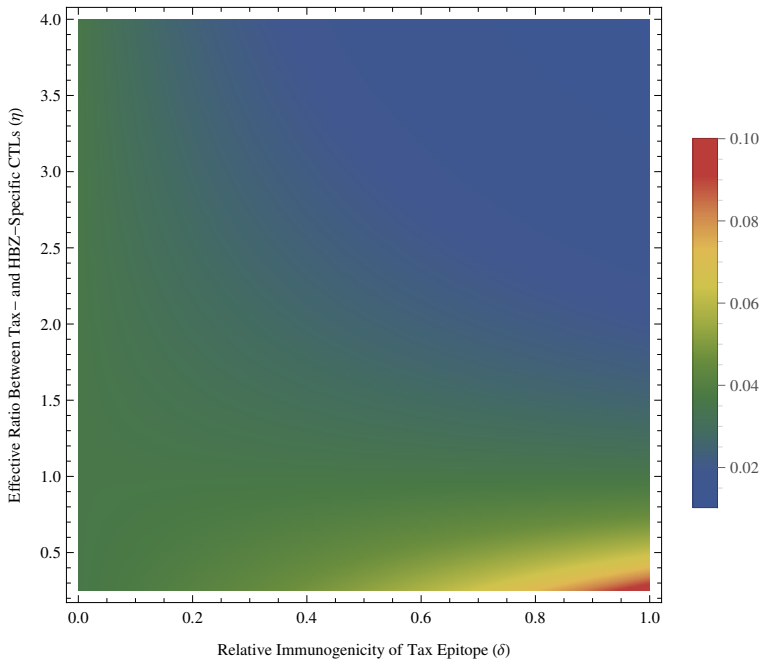


Figure 4.5: Heat map of PVL demonstrating how immunogenicity (δ) influences the magnitude of the PVL (v^*) in different ways depending on the effective ratio (η) between distinct CTL responses, thus illustrating the theoretical result derived in Theorem 4.3. Specifically, when $\eta < 1$, there is positive correlation between the PVL and δ. When $\eta = 1$, no correlation exists between the PVL and δ. Lastly, when $\eta > 1$, the PVL and δ display a negative correlation. The heat map shows that, as Tax immunogenicity (δ) increases, a skew in the effective ratio towards either Tax-specific CTLs (i.e. $\eta > 1$) or HBZ-specific CTLs (i.e. $\eta < 1$) can lead to a significant decrease or increase, respectively, in the size of the PVL. Parameter values are the same as those used in Figure 4.4, which have been taken from biologically realistic ranges as discussed in Table 3.1.

neutral point where both CTL types are effectively identical (i.e. $\eta = 1$), the magnitude of the PVL is unaffected by differences in immunogenicity between the two epitopes.

On the other hand, by taking vertical cross-sections, we can examine how the PVL changes with respect to differences in the effective ratio between the two CTL responses for a fixed immunogenicity. What we discover is that when the extent of immunogenicity of the Tax epitope is low (i.e. small δ), the variation between the two CTL types in terms of their effective ratio has little effect on the PVL. However, if the Tax epitope is highly immunogenic (i.e. large δ), then the degree of skew in the effective ratio can have a substantial impact on the magnitude of the PVL. Specifically, if the Tax epitope is highly immunogenic, then an increase in the effective ratio towards Tax-specific CTL properties is associated with a more dramatic decrease in the PVL. Because the effective ratio incorporates the efficiency of the two types of antigen-specific CTLs (i.e. γ, γ'), next we will briefly link the above results to what we have seen previously in the baseline model in terms of how varying the

efficiency of anti-HTLV-I immunity can affect the size of the PVL, and extend these ideas to take into account the role of immunogenicity.

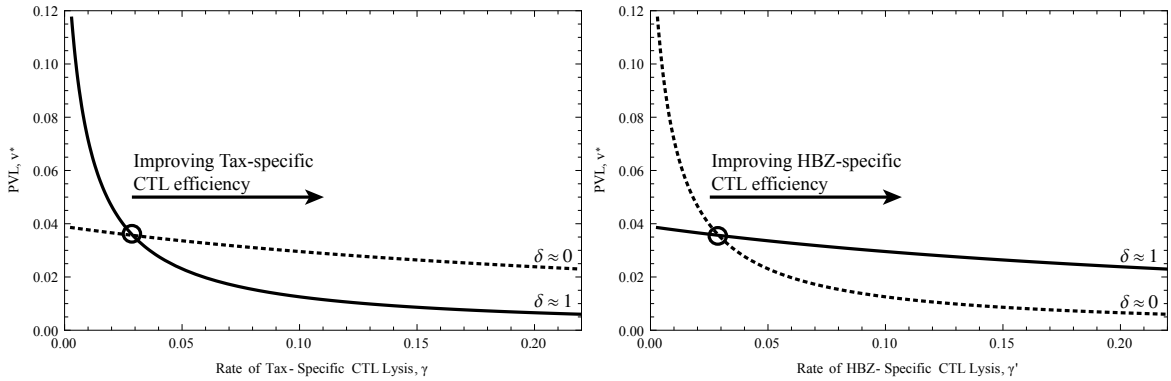
4.3.3 The Impact of the Efficiency of Each Antigen-Specific CTL Type Depends on Immunogenicity

One significant biological consequence from previous mathematical models of HTLV-I infection, including that of our baseline model (3.1), is that the measure of efficiency of the cellular immune response is not related to the frequency of CTLs, but rather to the rate at which CTLs lyse activated proviral cells. In the Tax/HBZ model (4.1), we consider two different populations of HTLV-I-specific CTLs, whose respective rates of CTL-mediated lysis are γ and γ'. As we have mentioned earlier, these two CTL responses are likely to be functionally distinct and hence would be expected to have a differential impact on the dynamics of the infection. What then happens if the efficiency of the cellular immune response is increased, for instance, by administering treatment that improves the rate of lysis of infected target cells? One question that arises immediately is: Does it matter which of the CTL responses, Tax-specific or HBZ-specific, is improved?

Using our Tax/HBZ model (4.1), we take a closer look at how the immunogenicity of Tax can affect the ability of each type of HTLV-I-specific CTL to control the PVL, $v^* = \big(u^* + y^*\big)/\big(x^* + u^* + y^*\big)$. Due to the widespread notion that the Tax epitope is immunodominant, a natural idea for reducing the PVL would be to increase the rate of lysis of CTLs specific to Tax, i.e. increase γ. However, a significant outcome of our investigations is that administering such a strategy of only improving the efficiency of Tax-specific CTLs may not necessarily be effective at lowering the PVL. The reason is that the degree of immunogenicity of each CTL epitope, captured by the parameter δ in model (4.1), is important in determining which of the antigen-specific CTL responses can be made more efficient to achieve a greater reduction in the PVL.

The efficiencies of Tax-specific CTLs and HBZ-specific CTLs characterised by the respective rates at which they eliminate actively infected target cells, γ and γ', with respect to the PVL, v^*, are plotted in Figure 4.6. Let us briefly describe Figure 4.6(a). The open circle $\bigcirc$ represents the magnitude of the PVL of a chronically infected individual with a given immune efficiency when $\gamma = \gamma'$, i.e. Tax-specific and HBZ-specific CTLs lyse actively infected cells at the same rate. The curves then represent the way in which the equilibrium PVL, v^*, varies (with all other parameters fixed) as the efficiency of Tax-specific CTL-mediated lysis, γ, changes for two separate cases: (1) Tax is highly immunogenic ($\delta \approx 1$, solid line), and (2) HBZ is highly immunogenic ($\delta \approx 0$, dashed line). In the former case, we observe, as expected, that the level of the PVL v^* experiences a more substantial change with respect

to the efficiency γ of Tax-specific CTLs when Tax is the principal source of stimulation for CTL proliferation ($\delta \approx 1$) than the case when HBZ primarily stimulates CTL expansion ($\delta \approx 0$), as shown in Figure 4.6(a). The converse holds for the efficiency of HBZ-specific CTLs, as depicted in Figure 4.6(b). In particular, if CTLs are largely stimulated by Tax ($\delta \approx 1$, solid line), then increasing the efficiency γ' of HBZ-specific CTLs has little impact on the level of the PVL at equilibrium. However, if CTLs are largely stimulated by HBZ ($\delta \approx 0$, dashed line), then an increase in HBZ-specific CTL efficiency γ' could lead to a significant reduction in the equilibrium PVL.



(a) Tax-specific CTL efficiency and the proviral load v^*. Solid line: $\delta \approx 1$. Dashed line: $\delta \approx 0$.

(b) HBZ-specific CTL efficiency and the proviral load v^*. Solid line: $\delta \approx 1$. Dashed line: $\delta \approx 0$.

Figure 4.6: The relative immunogenicity of either Tax or HBZ determines which of the antigen-specific CTL responses is more effective at controlling the magnitude of the proviral load at equilibrium, $v^* = \big(u^* + y^*\big)\big/\big(x^* + u^* + y^*\big)$. In both of the above figures, the solid black line represents the situation when Tax displays high immunogenicity, corresponding to large δ. Meanwhile, the dashed black line corresponds to the case when HBZ displays high immunogenicity and δ is small. The open circle $\bigcirc$ corresponds to the initial situation when the efficiencies of both types of CTL responses are equal, i.e. $\gamma = \gamma'$. (a) Boosting the efficiency of Tax-specific CTLs is effective at lowering the magnitude of the proviral load when δ is large (solid line), but not as effective when δ is small (dashed line). (b) The proviral load shows the converse effect with respect to efficiency of HBZ-specific CTLs. Namely, when δ is large (solid line), an increase in the efficiency of HBZ-specific CTLs has a relatively small impact on the proviral load as compared with the case when δ is small (dashed line), where more efficient HBZ-specific CTLs can have a strong effect on the magnitude of the proviral load.

Another way to look at Figures 4.6(a)–(b) is as follows. A high level of stimulation of Tax-specific CTLs is represented by the solid black lines (i.e. $\delta \approx 1$). The proviral load is seen to be dramatically affected by varying the efficiency γ of Tax-specific CTLs (Figure 4.6(a)), but is largely unaffected by changes to the efficiency γ' of HBZ-specific CTLs (Figure 4.6(b)).

The above theoretical result from our model (4.1) is biologically realistic as CTL quality

is believed to depend precisely on these two factors: (1) the removal of virus-infected cells by CTLs, and (2) CTL proliferation stimulated by the presence of antigen (Bangham et al., 2009; Wodarz et al., 2001). A biological implication of this result is that simply boosting the efficiency of any arbitrary anti-HTLV-I CTL response may not be beneficial in achieving a reduction in the proviral load. It may be important to first identify which pool of antigen-specific CTLs is being preferentially stimulated, and selectively boost the efficiency of that particular CTL response. For instance, if the Tax epitope is highly immunogenic (i.e. $\delta \approx 1$), then increasing the efficiency of the Tax-specific CTL response, say, by a certain factor, would result in a greater overall reduction in the proviral load rather than increasing the efficiency of the HBZ-specific response by the same factor. The identification of which CTL response would effect the greatest reduction in proviral load is clinically relevant and could have implications for the delivery of optimal treatment regimens.

4.4 Discussion of Main Results from the Tax/HBZ Model

In this chapter, we have developed a basic two-epitope model of persistent HTLV-I infection, called the Tax/HBZ model (4.1), to investigate the simultaneous expression of multiple viral epitopes in the presence of host immunity. The principal difference betweeen the Tax/HBZ model (4.1) and the baseline model (3.1) is the refinement of the HTLV-I-specific cellular immune response into two separate compartments, CTLs directed at Tax and CTLs directed at HBZ. Such a separation of distinct CTL responses allows us to study the role of anti-HTLV-I cellular immunity on the outcome of the infection in more detail. Our results help us to understand how host immunity reacts against infected cells that can stimulate distinct responses to different antigenic targets.

Below we summarise and discuss the main conclusions obtained from the Tax/HBZ model (4.1) which directly address the two motivating questions posed at the beginning of this chapter.

Q1. Which factors contribute to the prevalence of a certain CTL type when all target antigens can be recognised by host immunity?

▷ **The relative abundance of CTLs with respect to two different measures is dependent on two key factors: (i) relative immunogenicity towards a particular epitope, and (ii) the differences in the effective rates of expansion of each of the two CTL types.**

The parameter δ encapsulates the immunological meaning of immunogenicity of the CTL epitopes Tax and HBZ. Different values for δ correspond to the extent of stimulation of either Tax-specific CTLs or HBZ-specific CTLs. Meanwhile, the respective

effective CTL expansion rates for Tax- and HBZ-specific CTL responses are ϕ_{Tax} and ϕ_{HBZ}, and are determined by the growth and death rates of each CTL type.

From the Tax/HBZ model (4.1), we have derived two different measures of CTL prevalence: the relative abundance of each CTL type in the pool of effector cells, and the CTL-to-virus ratio, which is an indication of the abundance of CTLs in relation to the size of the proviral population. One of the main results that we have shown using the model equations is that both the structure of the CTL population, as well as the size of the effector pool with respect to either of the aforementioned measures, are wholly determined by the relative immunogenicities of the epitope targets (δ), and the ratio of the effective expansion rates between the two epitope-specific CTLs ($\phi_{\text{Tax}}/\phi_{\text{HBZ}}$). Moreover, we discovered that the efficiency of CTL-mediated lysis is not a contributing factor to the structure of the cellular immune response in terms of the relative abundances of each CTL pool.

Q2. Why can one CTL response have a greater impact on the proviral load than another? More broadly, what defines a 'good' immune response?

▷ **The extent of immunogenicity towards one epitope or another, and not only the expansion rates of the corresponding CTL response, determines how well epitope-specific cellular immune responses can control the proviral load and could possibly play a bigger role on other aspects of viral persistence and pathogenesis.**

In our previous analysis of the baseline model (3.1), we have demonstrated that the size of the HTLV-I-specific CTL response is by no means a definitive metric of its efficiency in eliminating virus-infected cells. Rather, we showed that the proviral load is affected by the efficiency of the CTL response in terms of the rate of CTL-mediated lysis. It can also be shown that the effective expansion rate (i.e. the proliferation and lifespan) of anti-viral CTLs is inversely correlated with the size of the proviral load.

In the Tax/HBZ model (4.1), we found that immunogenicity, in addition to the skew in the effective expansion rates and rates of cell-mediated lysis between the two distinct epitope-specific CTL responses (termed the effective ratio η), not only influences the prevalence of CTL effectors, but can also have a strong effect on the proviral load. Notably, we discovered that an efficient antigen-specific CTL response (i.e. one which is effective at maintaining a reduced PVL) is characterised by a fast rate of cell-mediated lysis and whose effective ratio (with respect to other CTL types) is skewed towards an epitope exhibiting a high degree of immunogenicity, both echoing and extending an earlier result derived from the baseline model (3.1).

4.5 Final Remarks on the Two-Epitope Tax/HBZ Model

The consideration of simultaneous expression of the HTLV-I antigens Tax and HBZ accompanied by multiple CTL responses adds a layer of complexity to the way we think about the host-virus interactions in HTLV-I infection. One important aspect that is taken into account in our Tax/HBZ model (4.1) is that an actively infected target cell displaying both viral proteins can interact simultaneously with multiple anti-HTLV-I cellular immune responses, Tax-specific or HBZ-specific.

The Tax/HBZ model (4.1) developed in this chapter captures a highly dynamic interaction among transcriptional latency, infected target cell activation accompanied by simultaneous viral protein expression, and antigen-specific anti-HTLV-I CTLs. The key difference between the baseline model (3.1) and the Tax/HBZ model (4.1) is the explicit incorporation of multiple epitopes and, correspondingly, multiple CTL responses, which allows us to explore a more complex and biologically representative within-host landscape that HTLV-I navigates amidst host immunity.

The study of our two-epitope Tax/HBZ model has illuminated a significant role for epitope immunogenicity in determining what constitutes an effective anti-viral cellular immune response with respect to reducing the magnitude of the proviral load. This observation prompts a deeper investigation into the link between epitope immunogenicity and other CTL-related properties. Indeed, in HTLV-I infection, an array of viral gene products can be targeted by host CTLs simultaneously (Parker et al., 1994). Although the analysis and results in this chapter have been developed for the scenario with two viral epitopes, the same methodology can be generalised to incorporate any number of distinct viral gene products, which we will investigate in the following chapter. Alternatively, the Tax/HBZ model (4.1) can be extended to take into account the phenomenon of antigenic variability, which is a mechanism occurring in a wide range of pathogenic invasions, including HTLV-I, whereby different variants of the same protein or epitope can exist at the same time. We will explore the implications of this latter idea in a subsequent chapter.

Chapter 5

Principles of Multiple Epitope Systems Without Antigenic Variation

In Chapter 4, we examined a simple extension of the baseline HTLV-I model (3.1) from Chapter 3 highlighting the role of simultaneous expression of two epitopes, namely Tax and HBZ. Because cellular immunity is highly specialised, this leads to the co-existence of two distinct CTL responses, each one specific to precisely one of the viral epitopes. A key feature of the two-epitope Tax/HBZ model (4.1) not present in the baseline model is the introduction of the concept of epitope immunogenicity, which refers to the immunological mechanism that despite the fact that both Tax and HBZ are displayed concurrently, for any given actively infected target cell, each of the viral gene products provides differential stimulation for the expansion of its cognate CTL type. The model predicted that in conjunction with differences between the effective CTL expansion rates for the two CTLs, the extent of immunogenicity was also important in determining the structure or relative abundance of each CTL type. Futhermore, the model predicted that while the efficiency of anti-HTLV-I CTL responses, in terms of the rate of CTL-mediated lysis, was negatively correlated with proviral load, a high degree of immunogenicity towards the particular epitope against which an efficient CTL response is directed played a pivotal role in effecting a reduction in the magnitude of the proviral load.

At this point, we recognise that there are a handful of ways in which we can extend the basic two-epitope Tax/HBZ model (4.1). There are two distinct approaches that we will develop in more detail in this and the forthcoming chapters, each motivated by zooming in on complementary aspects of the host-virus interaction. One way to generalise the Tax/HBZ model (4.1) is to incorporate any arbitrary number of epitopes, which we investigate thoroughly in this chapter. This approach raises deeper questions concerning epitope immunogenicity and the immunological phenomenon of immunodominance, and allows a

more careful examination of the properties of the host cellular immune response during persistent retroviral infection. In particular, we develop a framework to incorporate the diversity of anti-HTLV-I CTL responses and explore the factors that define distinct hierarchical structures in the CTL population, which can then differentially influence the control of the virus-infected cell pool. After laying down the theoretical groundwork, in the last part of this chapter, we demonstrate a proof of concept that our mathematical framework applied to a set of experimental data can be used for the estimation of parameters that may not be easily measurable. A second extension of the Tax/HBZ model (4.1) involves consideration of antigenic variability at each of the two epitopes, Tax and HBZ, which we will explore in the following chapter. We briefly mention here that the two approaches that we have outlined are not specific to HTLV-I infection and can be applied more generally to a wider range of host-pathogen interactions. Nevertheless, we will develop these ideas in the context of our HTLV-I modelling framework.

5.1 Motivation: What is Immunodominance and Why is it Important for us to Consider?

The consideration of multiple distinct antigen-specific CTL types in the HTLV-I system prompts an investigation along two main threads: (1) the structure of the CTL response, and (2) the concept of immunodominance. For the former, due to the diversity of CTLs, we raise the question as to whether the population of immune effectors self-organises into a CTL hierarchy in which various responses exist at different abundances and, if so, whether we can identify the key underlying factors that create and maintain such structure in the CTL pool. For the latter, we try to address an intriguing conundrum frequently observed in chronic viral infections, specifically that of CTL immunodominance, in the context of our HTLV-I model system. Intuition suggests to us that since there exists a wide array of different CTL epitopes that can all elicit an immune response, it might be expected that the pool of anti-viral CTLs would be diverse and consist of a wide and relatively uniform mix of epitope-specific responses. However, what is more commonly observed is *immunodominance*[1], a phenomenon occurring in a wide range of host-pathogen systems whereby the bulk of the immune response is directed against only one or two epitopes, despite there being a variety of CTL epitopes presented by the invading pathogen, most or even all of which can be recognised by host immunity (Akram and Inman, 2012; Frank, 2002). Meanwhile, the remaining subdominant immune responses are suppressed, even though they may be completely functional and in circulation in the peripheral blood.

[1] We refer the reader to Akram and Inman (2012) for a comprehensive review of immunodominance, with specific examples and detailed explanation taken in the context of infection by Influenza A and HIV-1.

With respect to chronic infection by HTLV-I, blood collected from seropositive individuals, regardless of clinical status, consistently shows that CTL responses restricted by a variety of host HLA Class I alleles can recognise a wide range of viral gene products (Goon et al., 2004a; MacNamara et al., 2010), such as the common retroviral proteins Gag, Env, Pol and Pro, as well as the HTLV-I regulatory proteins Tax, Rof, Tof, and the minus-strand HBZ (Goon et al., 2004a; Kannagi et al., 1991; Nagai et al., 2001; Pique et al., 2000). These sera studies typically show that in the majority of HTLV-I seropositive individuals, both asymptomatic carriers and patients with the neurological disease HAM/TSP, the HTLV-I-specific cellular immune response is principally directed towards epitopes in the Tax protein (Bangham, 2000b; Goon et al., 2004a; Jeffery et al., 1999; Kubota et al., 2007; Parker et al., 1994), thereby designating Tax as the immunodominant[2] epitope. Moreover, these Tax-specific CTLs often exist at frequencies higher than those found in most other chronic virus infections (Daenke et al., 1996), with the possible exception of HIV-1, and can even be present at abundances of 1–10% of all circulating $CD8^+$ effector CTLs among infected individuals with cellular immune responses restricted by the HLA-A*02 genotype, which has a particularly strong binding affinity to epitopes in the Tax protein (Bangham, 2000b; Kattan et al., 2009). Somewhat counter-intuitively, for a number of persistent infections, including HTLV-I, experimental studies have demonstrated that the immunodominant epitope may not necessarily be the optimal target for CTLs in terms of limiting the burden or degree of detriment caused by the virus, suggesting that further investigation is needed to evaluate what constitutes an effective, or protective, host immune response.

The key aspect that we would like to focus on in this chapter is to characterise heterogeneity in the host immune response and how it influences the outcome of infection. Ultimately, this requires us to build a better understanding of the phenomenon of immunodominance and the role that immunogenicity plays in establishing a hierarchy of epitope-specific CTL responses.

5.2 Extending the Tax/HBZ Model to Incorporate an Arbitrary Number of Distinct Epitopes

Model (4.1) can be straight-forwardly extended to incorporate any arbitrary number of epitopes. However, it is fundamental that we consider carefully the way in which proviral cells

[2]There are several different ways in which the concept of "immunodominance" of an antigen may be defined (Cook et al., 2013). It could refer to the antigen that is most frequently recognised by host immunity, thus highlighting the importance of immunogenicity of that particular antigen. Alternatively, and more commonly, an immunodominant antigen may refer to one that elicits the largest CTL response in terms of CTL abundance, either before or after correcting for the difference in protein lengths. In HTLV-I infection, it has been demonstrated that the Tax protein is immunodominant by any of the above definitions (Cook et al., 2013).

interact with a broader range of CTL types that are each specific to precisely one antigenic target. It is assumed that actively infected T-cells display their entire set of viral proteins, and that CTLs specific to each of the antigenic targets, or epitopes, exist. The extent to which one particular CTL response is stimulated over another is a function of the relative immunogenicity of the corresponding epitope. The key interactions between activated proviral T-cells expressing multiple epitopes simultaneously and the array of antigen-specific CTL responses is illustrated schematically in Figure 5.1, where we have sketched a representative scenario with five distinct epitopes.

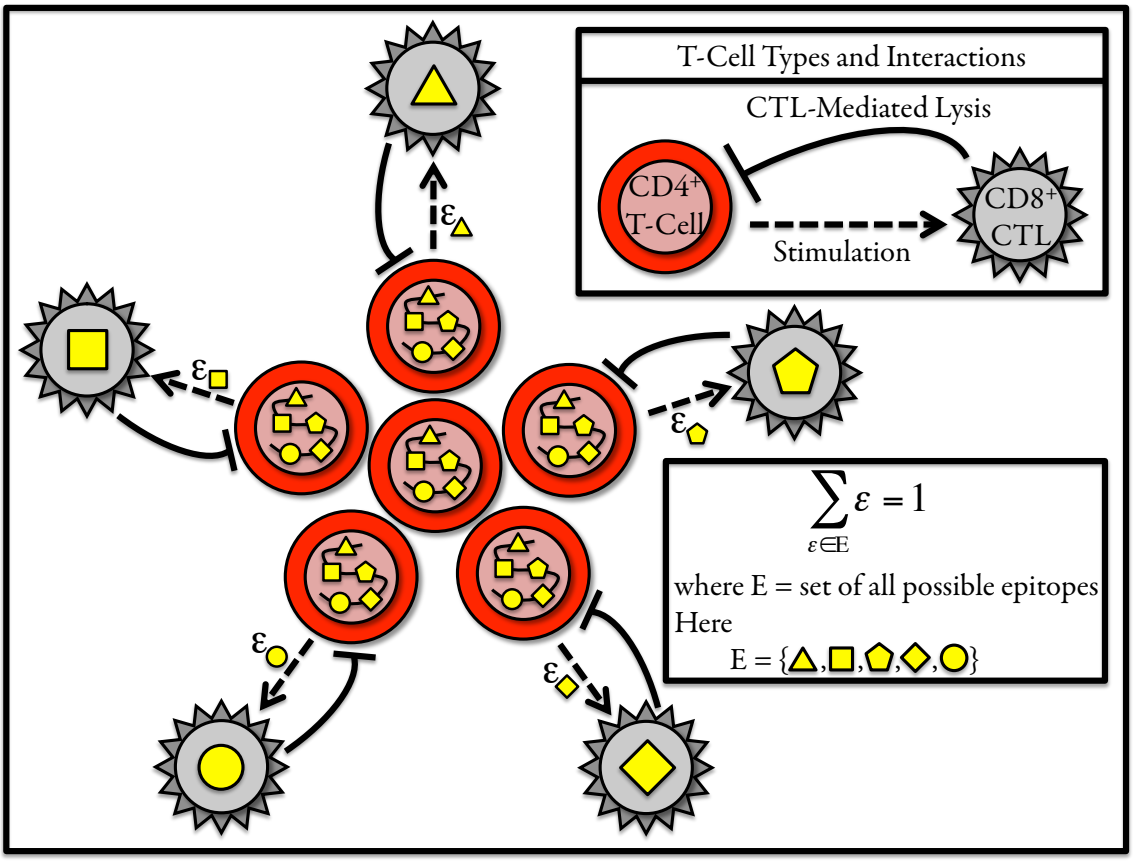


Figure 5.1: An illustration of the key interactions in a generalised multiple epitope model of HTLV-I infection. Actively infected target cells typically express a wide range of viral epitopes simultaneously (represented in the schematic figure above by the various yellow shapes). The extent to which each CTL type is stimulated relative to other CTL types depends on the immunogenicity of that particular epitope (i.e. the various 'ε' values corresponding to each of the yellow shapes). The result is that the host immune response is made up of a diverse range of epitope-specific CTLs. The relative abundance of each CTL type creates a dominance hierarchy for the anti-HTLV-I host immune response, with one response generally over-powering the rest of the responses. This phenomenon is known as immunodominance. Although a principal factor in establishing immunodominance, the concept of immunogenicity alone cannot explain immunodominance. In this chapter, we explore the various factors that contribute to immunodominance.

5.2.1 Formulating a General Multiple Epitope Model of HTLV-I Infection

Let us first remind ourselves of the notation for the T-cell classes under consideration, namely target cells and immune cells. The target cell population is comprised of $CD4^+$ helper T-cells and is subdivided into three compartments depending on which state it is in: healthy cells are those that are susceptible to infection, latently infected cells carry a transcriptionally silent copy of the provirus, and actively infected cells are activated proviral cells displaying the entire range of viral gene products. In the lattermost case, the simultaneous expression of all viral proteins then implies the existence of a variety of antigen-specific $CD8^+$ CTL responses that are able to recognise each of the individual epitopes from the various virus proteins. To formulate the above interactions mathematically, suppose that the HTLV-I provirus encodes $p \geq 1$ distinct viral epitopes, meaning that there are p epitope-specific CTL responses, which we will denote by $z_k(t)$, where $k = 1, \ldots, p$. Thus, we define the following compartments:

$x(t)$: number of healthy (uninfected) $CD4^+$ helper T-cells at time t,

$u(t)$: number of latently infected $CD4^+$ helper T-cells at time t,

$y(t)$: number of actively infected $CD4^+$ helper T-cells at time t,

$z_k(t)$: number of $CD8^+$ CTLs specific to epitope k at time t.

Combining the dynamics of the baseline model (3.1) and the Tax/HBZ model (4.1) in the presence of an arbitrary number of viral epitope targets allows us to construct the following generalised multiple epitope model for HTLV-I infection:

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \overbrace{\lambda}^{\text{T-cell production}} - \overbrace{\beta x y}^{\text{infectious transmission}} - \overbrace{\mu_1 x}^{\text{natural death}} \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= \overbrace{\beta x y}^{\text{infectious transmission}} + \overbrace{r y}^{\text{mitotic transmission}} - \overbrace{(\tau + \mu_2) u}^{\text{spontaneous activation and natural death}} \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \overbrace{\tau u}^{\text{spontaneous activation}} - y \overbrace{\sum_{k=1}^{p} \gamma_k z_k}^{\text{CTL lysis}} - \overbrace{\mu_3 y}^{\text{natural death}} \\
\frac{\mathrm{d}z_k}{\mathrm{d}t} &= \overbrace{\frac{\varepsilon_k}{\sum_{i=1}^{p} \varepsilon_i} \nu_k y}^{\text{CTL proliferation}} - \overbrace{\mu_{4,k} z_k}^{\text{natural death}}, \qquad \text{for} \quad k = 1, \ldots, p,
\end{aligned} \tag{5.1}$$

where $\varepsilon_k / \sum_{i=1}^{p} \varepsilon_i$ represents the relative immunogenicity of epitope k ($\varepsilon_k \geq 0 \ \forall k$).

Here we make the reasonable assumption that at least one of the $\varepsilon_k > 0$, otherwise no immune response would exist. We will prove in Section 5.2.2 that model (5.1) admits two non-negative equilibria: (1) The infection-free steady state, $P_0 = (x_H, 0, 0, \vec{0})$, where

$x_H = \lambda/\mu_1$ and $\vec{0} \in \mathbb{R}^p$, and (2) a positive, endemic steady state corresponding to chronic infection, $P^* = (x^*, u^*, y^*, \vec{z}^*)$, where $\vec{z}^* = (z_1^*, \ldots, z_p^*)$.

As with the previous models, we can derive the basic reproduction number for viral infection, R_0, representing the fitness of the virus, and it can be shown that the expression for R_0 in our multiple epitope model (5.1) is equal to the expression for R_0 in both the baseline model (3.1) and the two-epitope Tax/HBZ model (4.1). In particular, the basic reproduction number for viral infection for model (5.1) is defined as

$$R_0 = \frac{\tau}{\mu_3(\tau + \mu_2)}\Big(\beta x_H + r\Big), \quad \text{where } x_H = \frac{\lambda}{\mu_1}. \tag{5.2}$$

Remark 5.1. In the two-epitope Tax/HBZ model (4.1), we represented the relative immunogenicities of the Tax and HBZ epitopes using the parameters δ and $(1-\delta)$, respectively, where δ is a proportion lying between 0 and 1. Here, for the multiple epitope model (5.1), the relative immunogenicity of a particular epitope k, where $k \in \{1, \ldots, p\}$, is given by the fraction $\varepsilon_k / \sum_{i=1}^{p} \varepsilon_i$. Note that an equivalent description of the relative immunogenicity of epitope k, for $k = 1, \ldots, p$, would simply be the parameter ε_k subject to the condition $\sum_{k=1}^{p} \varepsilon_k = 1$. The quantity $\varepsilon_k / \sum_{i=1}^{p} \varepsilon_i$ is itself a proportion and so incorporates this requirement that the individual relative immunogenicities of each epitope sum up to unity.

Remark 5.2. Model (5.1) is indeed a direct generalisation of our previous models. Specifically, we point out that when $p = 1$, model (5.1) reduces to the baseline model (3.1). Moreover, when $p = 2$, setting $\delta = \varepsilon_1/(\varepsilon_1 + \varepsilon_2)$ yields that model (5.1) is equivalent to the Tax/HBZ model (4.1).

Remark 5.3. As with the two-epitope Tax/HBZ model (4.1), we conceptually separate the activation and effector phases of the cell-mediated adaptive immune response and represent them in the same way for the general multiple epitope model (5.1). In particular, although an actively infected target cell displays p distinct viral epitopes simultaneously, there is differential stimulation of each of the various CTL responses to proliferate during the activation phase. Meanwhile, we make the assumption that during the effector phase, each of the CTL types has an equal opportunity to eliminate any proviral cell. Thus, the factor expressing the relative immunogenicity of each epitope, given by the fraction $\varepsilon_k / \sum_{i=1}^{p} \varepsilon_i$ for epitope k, only appears in the term describing CTL proliferation and not in the terms describing CTL-mediated lysis.

5.2.2 Global Dynamical Behaviour of the Multiple Epitope Model

Here we use the methodology from the previous chapters and generalise the mathematical results therein to derive the global dynamics of the multiple epitope model (5.1) with an arbitrary number, $p \in \mathbb{N}$, of viral epitopes.

From the equations of the multiple epitope model (5.1), we observe that for any set of non-negative initial conditions, $x(t), u(t), y(t), z_k(t) \geq 0$ for all $t \geq 0$ and $k = 1, \ldots, p$. Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$, $\tilde{\nu} = \max_{k=1,\ldots,p} \nu_k$, and $\tilde{\mu} = \min_{k=1,\ldots,p} \mu_{4,k}$, and consider the closed, bounded region

$$\Gamma_p := \left\{ (x, u, y, z_k) \in \mathbb{R}_+^{p+3} : x \leq \frac{\lambda}{\mu_1},\ x + u + y \leq \frac{\lambda}{\mu - r},\ \sum_{k=1}^{p} z_k \leq \frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu - r)} \right\}.$$

Theorem 5.1. *Assume that $r < \mu$. Then, the set Γ_p defines a positively invariant region with respect to model* (5.1)*. All solutions are bounded for $t \geq 0$ and eventually enter Γ.*

The following theorem demonstrates that the basic reproduction number R_0 is a sharp threshold parameter that establishes the global dynamical behaviour of model (5.1).

Theorem 5.2 (Global Dynamics of the Multiple Epitope Model)**.**

1) *The infection-free equilibrium P_0 always exists. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma}_p$ and it is globally asymptotically stable in $\bar{\Gamma}_p$;*

2) *if $R_0 > 1$, then the infection-free equilibrium P_0 is unstable. In addition, a unique chronic infection equilibrium P^* exists in $\mathring{\Gamma}_p$ and it is globally asymptotically stable in $\mathring{\Gamma}_p$.*

The detailed proofs of Theorem 5.1 and Theorem 5.2 may be found in Appendix A.3.

5.2.3 Quantifying Measures of CTL Abundance: Establishing an Immunodominance Hierarchy

In the previous subsection, we have established the global dynamics of the general multiple epitope model (5.1), meaning that we have a thorough understanding of the qualitative behaviour of the model and can explore other relevant aspects of persistent HTLV-I infection that are represented by the model.

In the remainder of this chapter, we turn to the application of the model to the biological question posed as the motivation for our multiple epitope modelling approach, which is to better understand the complex interactions between HTLV-I proviral cells expressing an array of viral epitopes and an assortment of cellular immune effectors, namely $CD8^+$ CTLs, that are each highly specific to a given target antigen. From the equilibrium equations of the general multiple epitope model (5.1), we can derive two different expressions to quantify the size or influence of the CTL pool. Specifically, for each $k = 1, \ldots, p$, we define the relative abundance of CTL type k, and the ratio of CTL type k with respect to the number

of actively infected (or virus-expressing) target cells, as shown below in Equations (5.5) and (5.6), respectively.

From the equations in model (5.1), the frequency of each CTL type at equilibrium is

$$z_k^* = \frac{\varepsilon_k}{\sum_{i=1}^{p} \varepsilon_i} \phi_k y^*, \quad \text{for} \quad k = 1, \ldots, p, \tag{5.3}$$

where $\phi_k = \nu_k / \mu_{4,k}$ is the effective expansion rate, or growth/death ratio, of CTL type k. Pairwise comparison of Equation (5.3) for $k = 1, \ldots, p$ allows us to infer that any two CTL responses exist at equal equilibrium abundances when

$$\varepsilon_{k_1} \phi_{k_1} = \varepsilon_{k_2} \phi_{k_2}, \quad \text{for any} \quad k_1, k_2 \in \{1, \ldots, p\}. \tag{5.4}$$

For each $k = 1, \ldots, p$, we consider the relative abundance of CTL type k as a proportion of the total CTL population, which is given by the expression

$$\frac{z_k^*}{\sum_{j=1}^{p} z_j^*} = \frac{\varepsilon_k \phi_k}{\sum_{j=1}^{p} \varepsilon_j \phi_j}. \qquad \left[\text{Relative CTL Abundance}\right] \tag{5.5}$$

A simple re-arrangement of Equation (5.3) emphasises another related measure, the CTL to active virus ratio for each epitope-specific CTL response, which describes how many CTLs of a certain specificity there are scaled to the magnitude of the proviral load. For the multiple epitope model (5.1), the CTL-to-virus ratio for CTL type k, where $k = 1, \ldots, p$, is:

$$\frac{z_k^*}{y^*} = \frac{\varepsilon_k}{\sum_{i=1}^{p} \varepsilon_i} \phi_k. \qquad \left[\text{CTL-To-Virus Ratio}\right] \tag{5.6}$$

We observe that both of the above measures of CTL frequency can be expressed solely in terms of model parameters and so are independent of the density of infected cells, hence making them potentially useful indicators to compare the states of the immune responses among different HTLV-I-infected individuals. For the former, the relative abundance of each CTL type demonstrates how strongly each epitope elicits an immune response relative to all other epitopes, thus allowing us to construct a CTL frequency hierarchy listing epitopes from the most immunodominant to the least immunodominant. The ranking of CTL responses in terms of their respective relative abundances is known as the CTL *immunodominance hierarchy*[3]. Meanwhile, for the latter, the ratio of each CTL type to the size of the actively infected cell pool describes the influence that each of the various CTL responses has on the proviral target cell population.

[3]Note that the CTL immunodominance hierarchy differs from the ranking of CTL responses with respect to immunogenicity. The former is an ordering of CTL types with respect to their relative abundances in the CTL population, whereas the latter is an ordering of the corresponding epitopes in terms of how strongly each stimulates the proliferation of its respective effector CTL type.

The main point that we would like to highlight from the two measures of CTL abundance, Equations (5.5) and (5.6), is that the relative abundance of each CTL type k, where $k = 1, \ldots, p$, with respect to either of the two above CTL measures is determined only by two factors: the effective expansion rate (ϕ_k), and its respective immunogenicity (ϵ_k) relative to each of the other various epitopes. What this means is that an immunodominant CTL response is often one that both targets a highly immunogenic epitope and has a fast effective expansion rate. However, because neither immunogenicity nor the rate of effective expansion alone establishes CTL abundance, in our model framework it is possible to explain how the most immunogenic epitope need not elicit the largest response, or why the pool of CTLs that has the fastest rate of effective expansion need not necessarily be immunodominant. Notably, the above metrics describing the relative size of each epitope-specific CTL response are independent of the efficiency or rate of lysis of proviral target cells, which lends support to the proposed idea that the size of an individual's CTL response is not an accurate measure of its ability to reduce viral burden (Bangham, 2009; Bangham et al., 1996; Nowak and Bangham, 1996).

We briefly remark that the two CTL measures defined here are very closely related, as Equations (5.5) and (5.6) point out. Indeed, while the relative abundance of CTL type k is a weighted sum of both immunogenicity and effective expansion rate for CTL type k with respect to all other CTLs, the CTL to active virus ratio is determined by the effective expansion rate of CTL type k scaled by the relative immunogenicity of CTL type k without the need to consider weighted sums. Because our primary interest is to examine the hierarchy of the CTL response and how CTL population can impact the proviral load, we will focus our attention on quantifying CTLs according to the first metric, that is, the relative abundance of each CTL type.

Overall, it is important to identify such comparative metrics in order to construct a clearer picture as to why one immune response may be better or worse than another, especially when considered in conjunction with how a particular immune profile influences the traditional and other measures of viral detriment such as the proviral load and the activated proportion of proviral cells, both of which can exhibit a wide variance among infected hosts.

5.2.4 Constructing a Visual Representation of an Individual's CTL Response

One of the consequences of the simultaneous expression of multiple distinct viral epitopes by infected target cells is that the host immune response is correspondingly diverse due to its ability to recognise a wide range of foreign antigens. In our formulation for the

general multiple epitope model (5.1), we have assumed that proviral activation results in the expression of all viral epitopes and that the host cellular immune response is composed of the entire range of CTL types, each of which is specific to precisely one of the viral epitope targets.

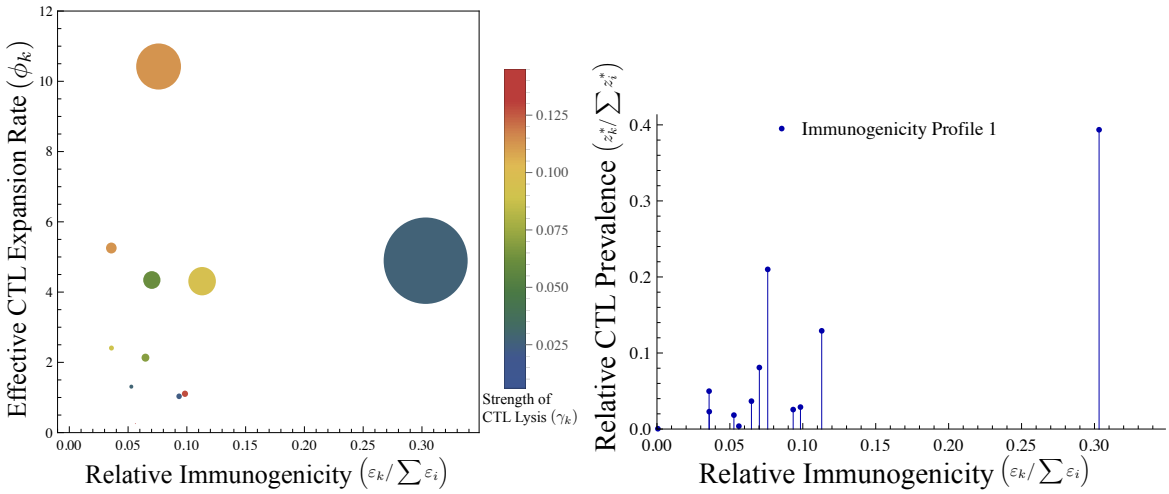
Each epitope-specific CTL type can be characterised by four main properties: (i) immunogenicity, (ii) effective expansion rate, (iii) efficiency of CTL-mediated lysis (effector function), and (iv) prevalence or relative abundance. In this section, we construct a visual representation of an individual's anti-HTLV-I cellular immune response that can capture these four attributes.

Figure 5.2 provides a graphical representation of the above concepts by considering the multiple epitope model (5.1) with p different types of epitope-specific CTL responses for $p = 12$, corresponding to 12 distinct epitopes on an activated proviral cell that can be recognised by host cellular immunity (see Remark 5.4 below for the reason why we have chosen $p = 12$ here). The effective expansion rates, relative immunogenicities, and efficiencies for the various CTL responses have been selected at random from physiologically meaningful ranges for each of the respective parameters from Table 3.1 as discussed in Chapter 3. We briefly recall that the relative abundance of each CTL type can be calculated from the effective expansion rate and the immunogenicity of the particular CTL type under study as in Equation (5.5). For reference, we refer the reader to Table 5.1, which lists the selected parameter values and properties for each of the CTLs represented in Figure 5.2.

In Figure 5.2(a), the attributes for each of the 12 distinct CTL responses are plotted in the form of a "colour bubble chart[4]" representation, with the position of each bubble on a two-dimensional "CTL fitness space", wherein the horizontal axis indicates the relative immunogenicity of the epitope, the vertical axis describes the effective expansion rate of the particular CTL type, and the colour of the bubble corresponds to its efficiency or rate of cell-mediated lysis according to a colour gradient. Lastly, the size of the bubble represents the relative abundance of that CTL type. Next, because one of the key attributes of our multiple epitope approach that differs from our earlier HTLV-I models is the incorporation of immunogenicity for an arbitrary number of epitopes in the dynamics of HTLV-I infection, to complement the bubble chart representation, in Figure 5.2(b) we illustrate the relative immunogenicity of each epitope corresponding to the CTL types in Figure 5.2(a) along with their relative abundances indicated by the height of the vertical lines using a "list plot[5]"

[4]The "colour bubble chart" representation is the combination of a standard bubble chart with an added dimension captured by colour according to a heat or temperature gradient, allowing us to illustrate four pieces of information at the same time. Namely, (i) position on the horizontal axis, (ii) height on the vertical axis, (iii) colour of the bubble, and (iv) size of the bubble.

[5]The "list plot" representation is a standard plot of points on a two-dimensional plane displaying two pieces of information. Namely, (i) position on the horizontal axis, and (ii) height on the vertical axis.



(a) CTL properties.

(b) CTLs all have equal immunogenicities.

Figure 5.2: Constructing a visual representation of the properties that define an individual's anti-viral cellular immune response. In this figure, each CTL type has a characteristic level of relative immunogenicity, effective expansion rate, and rate of lysis (i.e. CTL efficiency). (a) Bubble chart illustrating four properties for each of the p CTL types, where here we have taken $p = 12$: (i) relative immunogenicity (horizontal axis), (ii) effective expansion rate (vertical axis), (iii) efficiency of CTL-mediated lysis (colour of bubble according to a gradient), and (iv) CTL abundance (size of bubble). (b) Immunogenicity profile (IP) of the individual's cellular immune response as a list plot illustrating the relative immunogenicities (horizontal position of coordinate) and relative abundances (height of coordinate) corresponding to each CTL type. This visual representation focusses on the distribution of immunogenicities exhibited by the pool of CTL responses. The particular values for the four properties or attributes describing each of the 12 CTL types have been selected from biologically realistic parameter ranges as discussed previously and are summarised in Table 5.1.

representation, which defines what we call the *immunogenicity profile (IP)* of an individual's cellular immune response. Whereas the colour bubble chart representation provides a convenient way to visualise all four properties of each CTL type (Figure 5.2(a)), the list plot representation allows us to more easily see the distribution of relative immunogenicities towards each of the viral epitopes and thus emphasises the role of immunogenicity in creating structure in the CTL repertoire (Figure 5.2(b)). Collectively, these two representations allow us to compare how different immune configurations, either within a single individual or between separate individuals, can have a differential impact on the outcome of infection, namely factors related to viral detriment such as the proviral load or the active proportion of proviral cells, thereby providing us with a way in which we can visualise and subsequently evaluate the overall efficacy of an individual's immune response and explore the issue as to what constitutes an effective immune response as well as why one immune response may be better or worse than another.

The particular scenario illustrated in Figure 5.2 shows a typical choice of parameter values in our simulations, where we observe the trade-offs that exist among each of the CTL responses (see also the values listed in Table 5.1 for a quantitative comparison). For example, we notice that CTL type 8 has the fastest effective CTL expansion rate, has a relatively high rate of lysis, and exists at a proportionally high prevalence despite being only moderately immunogenic. By contrast, CTL type 1 is the most efficient at eliminating virus-infected target cells, but is nearly absent from the CTL population due to its poor immunogenicity combined with a relatively slow rate of expansion. Meanwhile, CTL type 12 is the most immunogenic as well as the most prevalent (hence we identify it as being the immunodominant CTL response), yet it is neither the fastest growing nor the most efficient. In the following section, we will explore in more detail how these attributes can combine together to create distinct CTL profiles that exhibit varied characteristics and exert differential pressure on the virus population.

Remark 5.4. MacNamara et al. (2010) created two ranked lists of HTLV-I epitopes according to their impact either on reducing the proviral or on preventing the onset of HTLV-I-associated disease. The ordering of these ranked lists were obtained from the analysis of experimental results that measured the strength of binding of individual virus proteins and its association with a reduction in proviral load and, independently, the relationship between binding strength and clinical status, i.e. whether there was a difference in binding strength of certain viral epitopes between asymptomatic carriers and patients with the neurological and inflammatory disease HAM/TSP. We have chosen $p = 12$ to coincide with the number of epitopes considered by MacNamara et al. (2010).

Remark 5.5. For a given epitope, the interaction of the MHC:peptide complex and the T-cell receptor (TCR) is involved in both the activation and effector phases of the cellular immune response. Due to being triggered by the same interaction, it is likely that there is a relationship between the parameters describing the effective CTL expansion rate and the rate of CTL-mediated lysis. Indeed, on the one hand, a high rate of CTL expansion results in a greater abundance of effector cells that can suppress viral burden. On the other hand, suppression of viral burden can also be brought about by a fast rate of CTL-mediated lysis. The observation that similar outcomes can potentially arise from the two aforementioned distinct processes suggests that a positive correlation may exist between the two parameters.

CTL Property	CTL Type ($p = 12$)											
	1	2	3	4	5	6	7	8	9	10	11	12
Relative Immunogenicity $\left(\frac{\varepsilon_k}{\sum_{i=1}^{p}\varepsilon_i}\right)$	0.000638	0.0357	0.0358	0.0528	0.0562	0.0648	0.0702	0.0759	0.0934	0.0984	0.113	0.303
Effective Expansion (ϕ_k)	2.379	5.254	2.408	1.308	0.255	2.134	4.344	10.421	1.034	1.106	4.310	4.894
CTL Efficiency (γ)	0.130	0.113	0.089	0.028	0.142	0.069	0.061	0.113	0.023	0.128	0.095	0.027
Relative Abundance $\left(\frac{z_k^*}{\sum_{j=1}^{p} z_j^*}\right)$	0.000403	0.0498	0.0229	0.0183	0.00380	0.0367	0.0809	0.210	0.0256	0.0289	0.129	0.394

Table 5.1: Properties for each of the $p = 12$ CTL types corresponding to Figure 5.2. Each CTL is defined by four attributes: (i) relative immunogenicity, (ii) effective expansion rate, (iii) efficiency of CTL-mediated lysis, and (iv) relative abundance. For each CTL property, the three largest values are highlighted in cyan (first), lime (second), and yellow (third) for emphasis. The values for each of the parameters have been selected from biologically realistic ranges as in Table 3.1.

5.3 Characterising the Anti-HTLV-I Cellular Immune Response Against Multiple Epitopes: Establishing a CTL Profile and Observing its Effect

The generalised multiple epitope model (5.1) provides us with a framework in which we can investigate in more detail the dynamics between a persistent virus expressing a range of different protein targets and a corresponding array of host immune responses specific to each of the viral protein targets. In this section, we develop the ideas of CTL hierarchy and immunodominance by utilising our visual representation of an individual's CTL repertoire described in the previous section, which characterises the principal properties of the cellular immune response, including the roles of epitope immunogenicities, effective CTL expansion rates, CTL efficiency, and CTL abundance, to explore how these properties influence the measures of viral burden such as the activated proportion of infected cells and the magnitude of the proviral load, both within the same individual as well as between different individuals.

5.3.1 The Immunogenicity Profile Alone Can Give Rise to an Array of Different CTL Population Structures, Even When Epitopes Preserve Relative Rank Order With Respect to Immunogenicity

A central feature of the present mathematical model (5.1) is the simultaneous expression of any number of viral epitopes accompanying proviral activation of infected target cells. Biologically speaking, it is known that host cellular immunity has evolved so as to create a diverse pool of CTLs that is able to recognise an extensive range of foreign antigens (Parham, 2005), and so in our model, we assume that CTLs targeting the full set of epitopes exist. The relative immunogenicity of each viral epitope is characterised by its ability to stimulate its cognate epitope-specific CTL response, and is determined by factors such as the strength of binding affinities to distinct viral peptide sequences from the particular epitope, the number of viral epitopes present, as well as the relative immunogenicities of all other epitopes. The collection of relative immunogenicities of all epitopes (which are individual proportions that sum up to unity) forms an immunogenicity profile with each epitope-specific CTL response ranked in order of increasing relative immunogenicity.

Within a single individual, fluctuations in the abundances of various epitope-specific CTLs are known to occur over time (Daenke et al., 1996; Parker et al., 1994). However, two of the defining properties of a given CTL type, namely the effective CTL expansion rate and the rate of CTL-mediated lysis, are intrinsic characteristics and therefore would not be expected to change throughout the course of chronic infection. How, then, can we reconcile the observed variation in the immunodominance hierarchy or structure of the cellular immune response with the stability of intrinsic CTL properties? Earlier in this

chapter, in Section 5.2.3, we have quantified the relative abundance of each CTL type according to Equation (5.5), demonstrating that it depends not only on the effective rate of CTL expansion (i.e. ϕ_k for each $k = 1, \ldots, p$), but also on the immunogenicity of the particular epitope (i.e. ε_k for $k = 1, \ldots, p$). Even if the intrinsic rate of CTL proliferation for each epitope-specific CTL type does not vary from one point in time to the next, a change in the structure of the CTL population can arise from variation in the immunogenicity profile, which is likely to fluctuate over the course of persistent infection. Such a scenario may arise, for instance, following the emergence of a previously unrecognised epitope or, alternatively, following the loss of recognition of a previously seen epitope over time[6].

To explore this aspect, let us consider a variety of different immunogenicity profiles for a given individual. These various immunogenicity profiles are generated by allocating specific values for the relative immunogenicities of each CTL response randomly from a uniform distribution between 0 and 1, all the while maintaining the same rank order of CTL epitopes with respect to immunogenicity[7]. The rank order of CTLs is itself assigned arbitrarily, then fixed thereafter. Because we are focussing on a single host, we assume that the other properties that define each CTL response, namely, the characteristic effective CTL expansion rate and rate of cell-mediated lysis, are all held constant. The parameters for CTL expansion and CTL efficiency have been selected randomly within biologically realistic ranges as described in Table 3.1 and are the same as the values listed in Table 5.1.

Figure 5.3 illustrates the properties of the cellular immune response from this same individual with variations only in the immunogenicity profiles. We consider five different immunogenicity profiles, which we label IP-1 to IP-4, plus IP-E corresponding to the situation in which all of the epitopes are equally immunogenic. These five immunogenicity

[6]In an experimental study, Parker et al. (1994) observed differential epitope recognition by the same HTLV-I-infected individual at different time points. For example, freshly isolated $CD8^+$ cells from patient HM at an initial time point showed recognition of four HTLV-I epitopes from the N-terminus region of Tax (Tax 12–19, XN3, XN4, and XN11) and one epitope, X4, from the middle of the molecule; meanwhile, a fresh sample taken three months later revealed the loss of recognition of XN11 and X4, but recorded the recognition of an additional epitope, X7, from the middle of Tax.

[7]Despite the mutability of individual immunogenicity values for each epitope, the CTL rank order with respect to the immunogenicities of HTLV-I epitopes is believed to remain relatively stable over time meaning that, for example, if some epitope A is more immunogenic than another epitope B, then it will generally continue to be so. There are several lines of evidence from experimental studies that lend support to this idea. First, the viral protein Tax, which is expressed early on in the infection, is the most recognised HTLV-I antigen and elicits an immunodominant anti-HTLV-I CTL response which is detected in almost all HTLV-I seropositive individuals without malignant disease, both asymptomatic carriers and HAM/TSP patients, and is often present at high frequencies even after years of persistent infection (Daenke et al., 1996; Goon et al., 2004a; Parker et al., 1992). Second, in the study by Parker et al. (1994), the majority of HTLV-I epitopes that were recognised at an intial time point were again recorded in a later time point from freshly isolated $CD8^+$ CTLs from three HTLV-I-infected individuals (two asymptomatic carriers and one patient with HAM/TSP). Third, it has been shown that virus variants that escape CTL recognition do not reach fixation and thus a particular epitope-specific CTL would continue to receive stimulation from its cognate antigen to proliferate (Niewiesk et al., 1995).

profiles are meant to demonstrate the typical kinds of patterns that can be observed in the CTL population. The specific values for the relative immunogenicities and relative abundances corresponding to each of the CTL types with respect to the five distinct, randomly generated immunogenicity profiles (IP-1, IP-2, IP-3, IP-4, and IP-E) can be found in Table 5.2. Because both epitope immunogenicity and the effective expansion rates of cognate CTLs determine CTL abundance according to Equation (5.5), it is unsurprising that a change in the distribution of relative epitope immunogenicities would change the prevalence of each CTL type. However, we point out that such a re-structuring of the CTL population can be brought about even though the rank order of the set of CTL responses is kept the same. Significantly, the immunogenicity profile alone can change the hierarchy of the CTL population and can even result in replacement of the immunodominant epitope — in other words, immunodominance can change.

To highlight this point from our current example, we compare the immunogenicity profiles IP-1 with IP-2 in the same individual as in Figure 5.3. Under immunogenicity profile IP-1, we observe that the cellular immune response is dominated by a single CTL type 12 targeting a highly immunogenic epitope, which at 39% prevalence in the effector pool, is almost twice the abundance as the next largest effector type, CTL type 8, targeting a much less immunogenic epitope and comprising roughly 21% of all CTLs. In contrast, under immunogenicity profile IP-2, the bulk of the immune response is made up of three effector types which, listed in order of prevalence, are CTL types 8, 12, and 11, although all three immune responses exist at fairly similar abundances, approximately 25%, 22%, and 18%, respectively. In addition to exhibiting significantly different CTL population structure, we also notice that the immunodominant response (which we define as being the one existing at the highest prevalance), differs between the two immunogenicity profiles. Specifically, for IP-1, CTL type 12 not only targets the most immunogenic epitope, but is also the immunodominant response, whereas for IP-2, CTL type 8 displays immunodominance, yet its target antigen is only moderately immunogenic. In the latter situation, the most immunogenic epitope 12 instead elicits a subdominant immune response by its corresponding CTL type 12. The other immunogenicity profiles each display their own unique set of characteristics as well. For instance, whereas the immunodominance of CTL type 8 (as well as the order of the second and third most abundant responses being exhibited by CTL types 12 and 11, respectively) is preserved under both IP-3 and IP-4, the two immunogenicity profiles differ to each other in terms of the individual values for the relative immunogenicities of the epitopes. In particular, we notice that the epitopes under IP-3 are more clustered in their relative immunogenicities than under IP-4. Similarly, under IP-E, CTL type 8 is also immunodominant, even though in this case, all epitopes are equally immunogenic. In the next

subsection, we will examine how these distinct immunogenicity profiles can differentially impact control of the infected target cell pool.

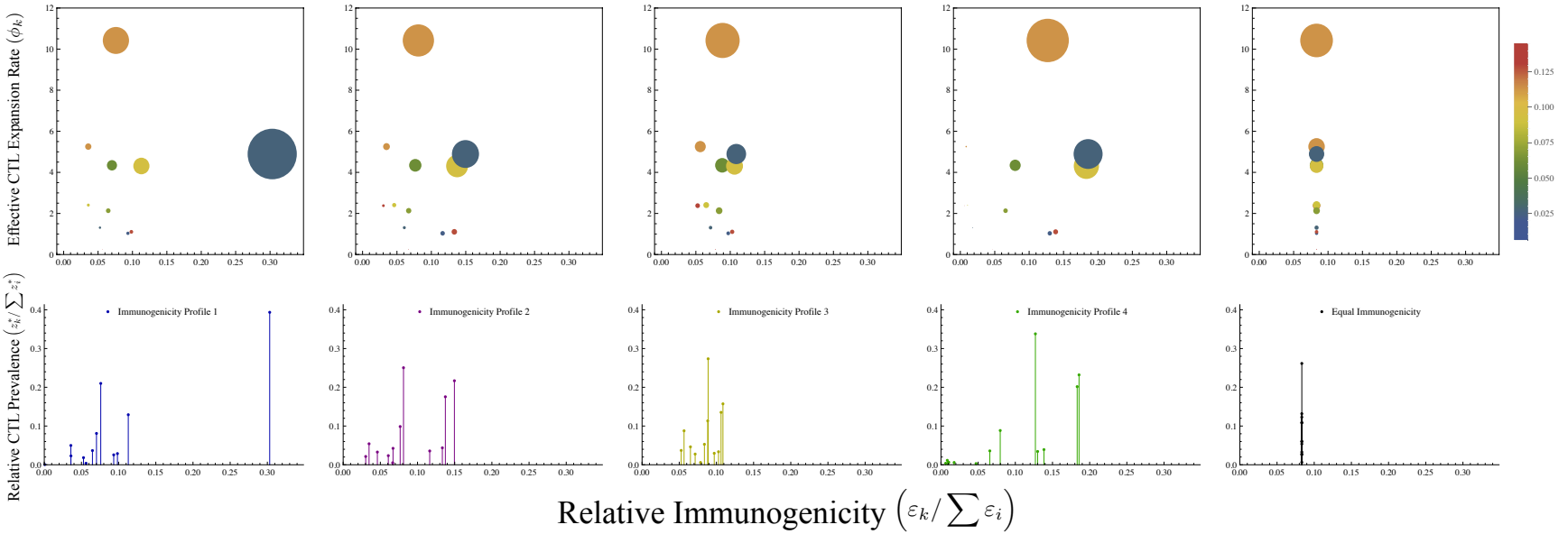


Figure 5.3: The structure of the CTL response for various different immunogenicity profiles within the same individual according to the multiple epitope model (5.1). The figure uses the two graphical representations introduced in Section 5.2.4 for visualising the properties of the CTL repertoire for each of five immunogenicity profiles, where each column corresponds to one of the immunogenicity profiles (c.f. Figure 5.2). From left to right: IP-1, IP-2, IP-3, IP-4, and IP-E. Top Row: The 'colour bubble' representations under the various immunogenicity profiles illustrate the following CTL attributes: (i) relative immunogenicity (horizontal axis), (ii) effective expansion rate (vertical axis), (iii) efficiency of CTL-mediated lysis (colour of bubble according to a gradient), and (iv) relative abundance (size of bubble). Bottom Row: The corresponding 'list plot' representations for each of the five immunogenicity profiles highlight the distinction of relative immunogenicities towards the various viral epitopes along with the abundance of their respective cognate CTLs. By comparing the visual representations for IP-1 through to IP-E (i.e. from left to right column-wise), our model demonstrates that the differences in immunogenicity profile alone can account for the variation seen in the structure and immunodominance hierarchy of the anti-HTLV-I CTL response over the course of chronic infection. This is due to the observation that the relative abundance of each CTL type is dependent on relative immunogenicity in addition to effective CTL expansion rate. The values for the relative immunogenicities and abundances of each CTL type for the immunogenicity profiles considered here are listed for reference in Table 5.2.

Immunogenicity Profile	CTL Property	CTL Type											
		1	2	3	4	5	6	7	8	9	10	11	12
IP-1	Immunogen.	0.000638	0.0357	0.0358	0.0528	0.0562	0.0648	0.0702	0.0759	0.0934	0.0984	0.113	0.303
	Abundance	0.000403	0.0498	0.0229	0.0183	0.00380	0.0367	0.0809	0.210	0.0256	0.0289	0.129	0.394
IP-2	Immunogen.	0.0303	0.0348	0.0460	0.0607	0.0666	0.0670	0.0766	0.0811	0.116	0.133	0.137	0.150
	Abundance	0.0214	0.0542	0.0328	0.0235	0.00504	0.0424	0.0986	0.250	0.0356	0.0437	0.175	0.217
IP-3	Immunogen.	0.0526	0.0565	0.0651	0.0714	0.07867	0.0838	0.0883	0.0888	0.0971	0.103	0.106	0.109
	Abundance	0.0370	0.0877	0.0463	0.0276	0.00593	0.0529	0.113	0.274	0.0296	0.0336	0.135	0.157
IP-4	Immunogen.	0.0063	0.0085	0.0105	0.0177	0.0472	0.0657	0.0797	0.127	0.13	0.139	0.183	0.186
	Abundance	0.00383	0.0114	0.00646	0.00591	0.00308	0.0358	0.0884	0.338	0.0343	0.0391	0.202	0.232
IP-E	Immunogen.	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12
	Abundance	0.0597	0.132	0.0604	0.0328	0.00641	0.0535	0.109	0.262	0.0259	0.0278	0.108	0.123

Table 5.2: Relative immunogenicities and CTL composition for each of the five immunogenicity profiles (IP-1, IP-2, IP-3, IP-4, and IP-E) corresponding to Figure 5.3. For each immunogenicity profile, the three most prevalent CTL responses in descending order (and their corresponding relative immunogenicities) are highlighted in cyan (immunodominant), lime (second-most abundant), and yellow (third-most abundant), respectively.

5.3.2 Immunogenicity Profile Can Have a Significant Impact on Viral Burden Because it Affects How the Rate of Spontaneous Viral Expression Differentially Manifests Itself

The effector function of the cellular immune response is responsible for eliminating virus-infected cells and reducing the detriment due to infection. In Chapter 3, we considered two measures of viral burden that could be investigated using the baseline HTLV-I model (3.1), namely the proviral load, which we defined as the proportion of total target cells that carry the virus, and the active proportion of proviral cells, which represents the fraction of infected cells that are activated and therefore subject to immune surveillance and selection.

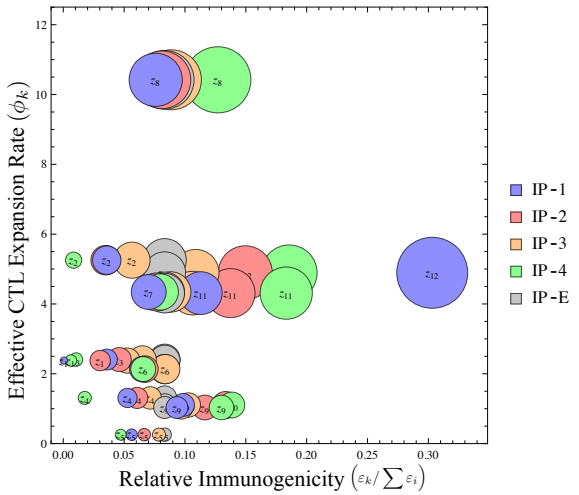
Our multiple epitope modelling approach incorporates an array of distinct CTL responses, each of which has its own set of characteristic properties. Differences in the structure of the effector population would emphasise different characteristics of the cellular immune response, which might then have a differential impact on the two measures of viral burden. We have seen previously that the immunogenicity profile within a single individual can give rise to substantially different CTL repertoires. In this subsection, we explore how these differing CTL repertoires impact the virus-infected target cell population.

Figure 5.4 illustrates many of the attributes of the CTL response for each of the five different immunogenicity profiles under consideration and the resultant measures of viral burden or detriment. In particular, Figures 5.4(a)–(b) show bubble chart and list plot representations for each of the CTL responses in a single individual for the five different immunogenicity profiles (IP-1, IP-2, IP-3, IP-4, and IP-E) considered in Subsection 5.3.1. Note that in the bubble chart visualisation here, unlike for the previous bubble chart representation for a given immunogenicity profile, we have not utilised a colour gradient for the efficiency of each CTL response since we have chosen instead to use different colours to distinguish between the various immunogenicity profiles (compare Figure 5.4(a) with Figure 5.2(a)). These alternative visual representations allow us to compare how the distribution of the CTL response lies in 'CTL fitness space' or along the 'immunogenicity space' for each of the five different immunogenicity profiles. Meanwhile, Figure 5.4(c) depicts the parametric curves describing the relationship between the proviral load and the active proportion of proviral cells, as governed by the rate τ of spontaneous proviral expression as the independent parametric variable, according to the multiple epitope model (5.1) under each set of specified parameter values corresponding to a particular immunogenicity profile[8]. The colours each represent a particular immunogenicity profile as indicated. We observe that the immune response of an individual can exert differential control of viral burden

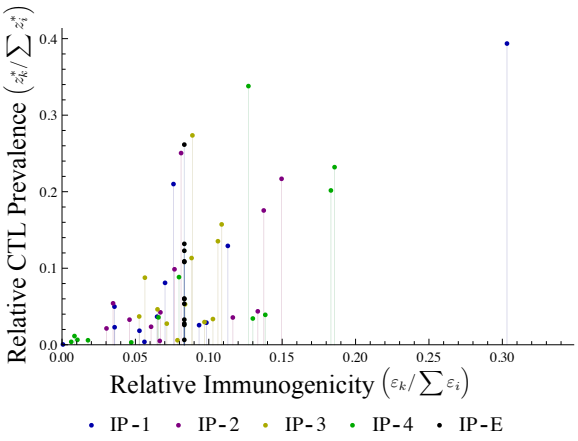
[8]Recall that we introduced this way of visualising the two measures of viral burden in our investigation of the baseline model (3.1) (c.f. Figure 3.11).

with respect to the magnitude of the proviral load and the activated proportion of proviral target cells depending on the immunogenicity profile, even when all other factors governing cellular immunity are held constant. This is because differing immunogenicity profiles alter the CTL population structure and therefore change the way in which each individual CTL type interacts with the virus-infected target cell pool.

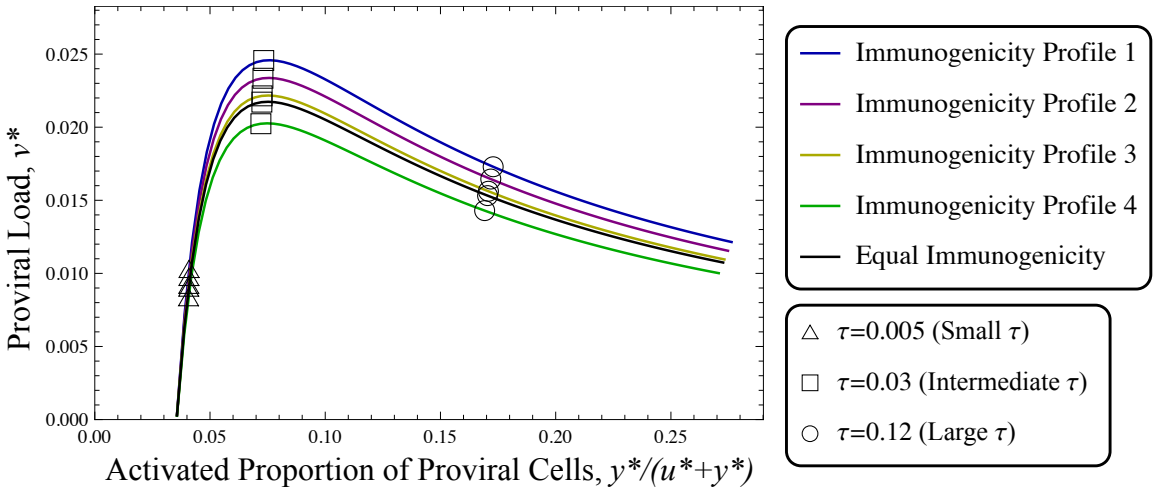
Moreover, we discovered that for different immunogenicity profiles, a given rate of spontaneous viral expression, which is represented by a fixed value of the parameter τ in model (5.1) and which determines the fitness of the virus in terms of its basic reproduction number R_0, differentially manifests itself with respect to the two measures of viral burden, i.e. the proviral load and the active proportion of proviral cells. Specifically, the same rate of spontaneous proviral expression can be associated with a higher proviral load and higher activated proportion under one immunogenicity profile, whilst manifesting as a lower proviral load and lower activated proportion under another immunogenicity profile. For instance, if we compare the two measures of viral burden for immunogenicity profiles IP-1 versus IP-4 in Figure 5.4(c), we observe a higher proviral load and activated proportion of proviral cells for the former immunogenicity profile (IP-1) as opposed to the latter one (IP-4) for the same fixed value of τ (whether or not τ is small, intermediate, or large). Similar qualitative comparisons can be made regarding the other immunogenicity profiles. The quantities for the two measures of viral burden for fixed values of τ under each of the five different immunogenicity profiles are shown for reference in Table 5.3.



(a) Bubble chart indicating the position of each CTL type in CTL fitness space.



(b) CTL immunogenicity profile and dominance hierarchy.



(c) Proviral load and active proportion as parametric functions of τ, the rate of spontaneous proviral activation and viral protein expression.

Figure 5.4: Exploring the role of CTL heterogeneity arising from differences in immunogenicity profiles for the same individual on the outcome of infection with respect to two measures of viral burden according to the multiple epitope model (5.1). Each of the following colours correspond to a single immunogenicity profile (unless otherwise indicated): IP-1 (blue), IP-2 (violet), IP-3 (yellow), IP-4 (green), IP-E (grey/black). (a) Bubble chart representation showing the CTL 'fitness space' and relative abundance of each CTL type for the five different immunogenicity profiles. (b) List plot representation showing the relative immunogenicities and prevalences of each CTL type for the five different immunogenicity profiles. (c) The curves corresponding to each of the immunogenicity profiles depicting the relationship between the two measures of viral burden, the proviral load, v^*, and the active proportion of proviral cells, $y^*/(u^*+y^*)$, as parametric functions of the rate of spontaneous infected target cell activation, τ, representing the fitness of the virus. Each of the curves is obtained by numerically solving model (5.1) at equilibrium with τ as the bifurcation parameter, and plotting the relationship between v^* and $y^*/(u^*+y^*)$ (note that we previously introduced this graphical representation for the baseline model (3.1), c.f. Figure 3.11). Selected values for τ, whether small ($\tau = 0.005$), intermediate ($\tau = 0.030$), or high ($\tau = 0.120$), demonstrate that the manifestation of the virus, for some fixed choice of τ, can differ depending on the immunogenicity profile, even when all other attributes are held constant.

Immunogenicity Profile	Proviral Load (PVL)			Active Proportion		
	Small τ ($\tau = 0.005$)	Intermediate τ ($\tau = 0.03$)	Large τ ($\tau = 0.12$)	Small τ ($\tau = 0.005$)	Intermediate τ ($\tau = 0.03$)	Large τ ($\tau = 0.12$)
IP-1	0.0103	0.0246	0.0173	0.0409	0.0733	0.173
IP-2	0.00976	0.0233	0.0165	0.0408	0.0730	0.172
IP-3	0.00922	0.0221	0.0156	0.0408	0.0726	0.171
IP-4	0.00837	0.0202	0.0143	0.0407	0.0721	0.169
IP-E	0.00903	0.0217	0.0153	0.0408	0.0725	0.170

Table 5.3: Magnitude of the proviral load and the active proportion of proviral cells at equilbrium for selected levels of viral fitness (in terms of the extent of spontaneous proviral activation τ which determines the basic reproduction number R_0 of the infection) for each of the five immunogenicity profiles corresponding to the markers in Figure 5.4(c). The top three immunogenicity profiles which give rise to the highest PVL and activated proportions of proviral cells for the same respective rates of spontaneous infected target cells activation τ are highlighted in cyan (first), lime (second), and yellow (third) for comparison.

5.3.3 Comparison of Cellular-Mediated Immunity Between Individuals: What Determines an Effective CTL Response?

In this subsection, we extend the analysis and simulation of the previous sections, which examined what happens within a single infected host, and explore how we can compare the various characteristics of the anti-HTLV-I cellular immune response between different infected hosts in the context of our multiple epitope framework. Figures 5.5 and 5.6 illustrate several distinct CTL profiles and their impact on viral burden by varying the immunogenicities and effective expansion rates of p different types of epitope-specific CTL responses across five different sample data sets which have been simulated using the equations for the multiple epitope model (5.1) when $p = 12$. These five different sets of sample data are meant to depict the typical kinds of behaviour that can arise. We refer to each of the five sets of sample data as SD-1, SD-2, SD-3, SD-4 and SD-C, the last of which represents the 'control' sample in which all CTL responses have the same effective expansion rates and immunogenicities (hence exist at equal abundances according to Equation (5.5)). For a given sample, each of the 12 different epitope-specific CTL responses have been assigned effective expansion rates and relative immunogenicities randomly from physiologically realistic parameter ranges as discussed for the baseline model (3.1). The CTL responses are ordered with respect to increasing immunogenicity so that $\varepsilon_{k_1} \leq \varepsilon_{k_2}$ for all $k_1 \leq k_2$, where $k_1, k_2 \in \{1, \ldots, p\}$. Lastly, for each $k = 1, \ldots, p$, the efficiency of cell-mediated lysis for CTL type k is assigned at random within a biologically relevant parameter range, and is the same across all sample data sets, e.g. CTL type 1 has the same rate of lysis in sample data SD-1 as in sample data SD-2, and so on.

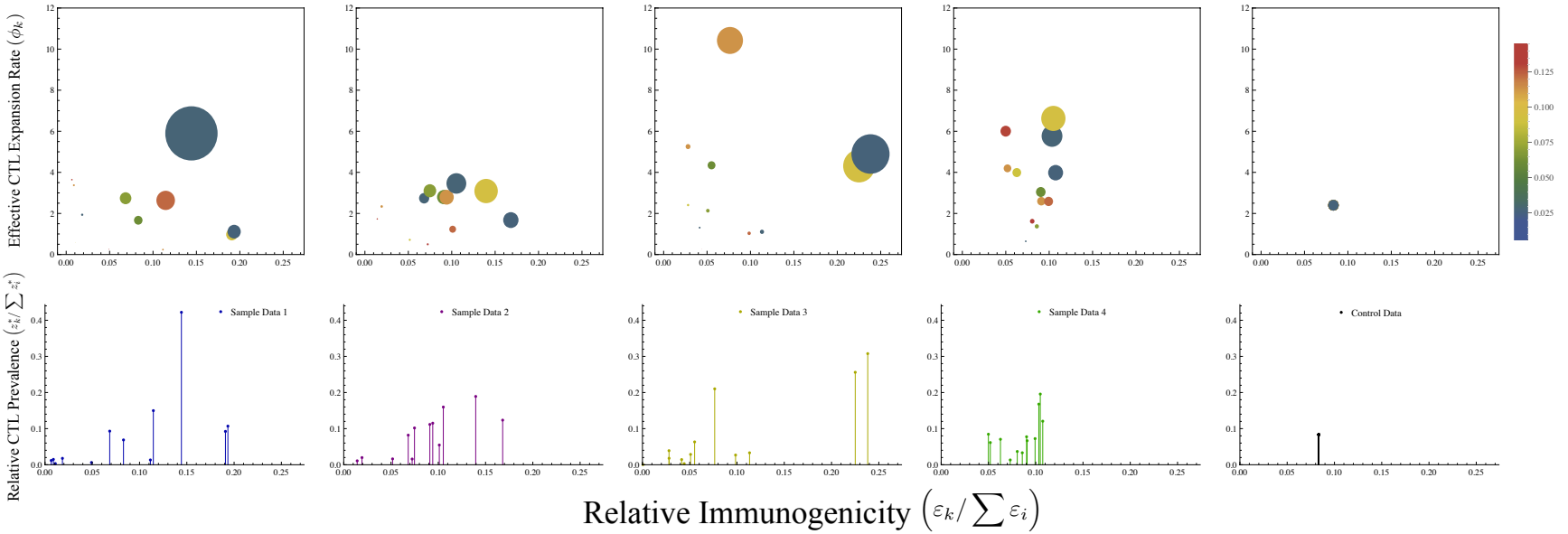
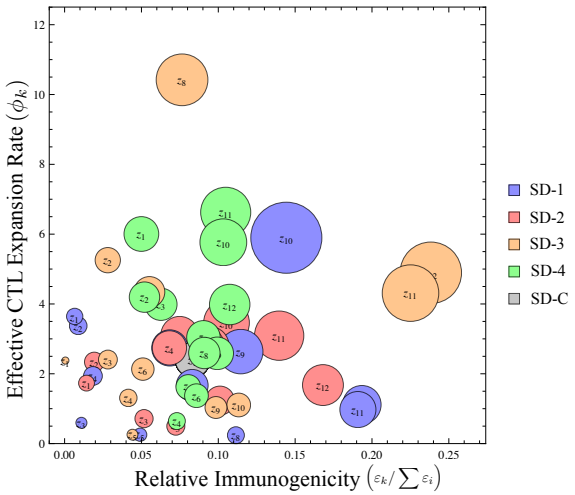
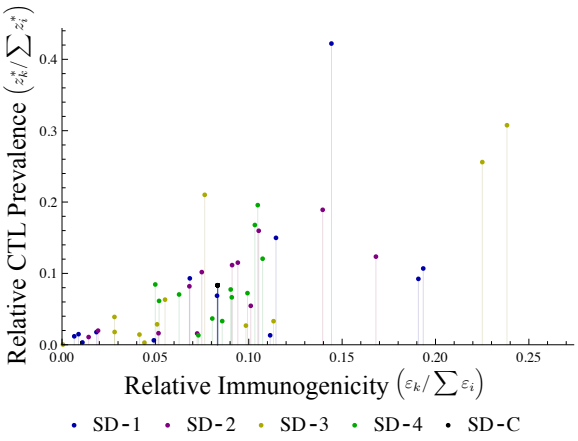


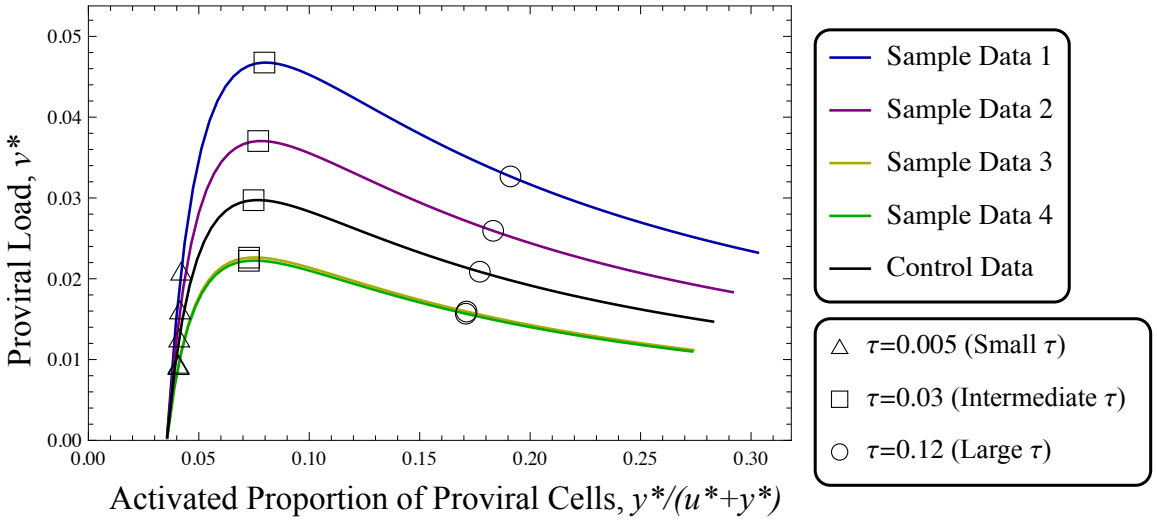
Figure 5.5: CTL properties and immunogenicity profiles for the immune responses of five different individuals. Each column shows the various CTL properties for the five sample data sets considered here, namely SD-1, SD-2, SD-3, SD-4, and SD-C, using the graphical visualisation representations introduced in Section 5.2.4 (c.f. Figure 5.2). Top Row: The 'colour bubble chart' representations for each sample data set illustrate four CTL attributes: (i) relative immunogenicity (horizontal axis), (ii) effective expansion rate (vertical axis), (iii) efficiency of CTL-mediated lysis (colour of bubble), and (iv) relative abundance (size of bubble). Bottom Row: The corresponding 'list plot' representations for each sample data set emphasise the relationship between the immunogenicity profile and the relative abundances of each CTL type. Parameter values and the corresponding relative abundances of effector T-cells defining the CTL profiles for each of the individuals (referred to, collectively, as 'sample data') are shown for reference in Table 5.4).



(a) Bubble chart indicating the position of each CTL type in CTL fitness space.



(b) Immunogenicity profile of each CTL type.



(c) Proviral load and active proportion as parametric functions of τ, the rate of spontaneous proviral activation and viral protein expression.

Figure 5.6: CTL dominance hierarchy and its effect on the virus population across five different sample individuals according to the multiple epitope model (5.1). Each of the following colours correspond to a particular sample data set (unless otherwise indicated): SD-1 (blue), SD-2 (violet), SD-3 (yellow), SD-4 (green), SD-C (grey/black). The latter-most set (SD-C) represents the control data set in which all CTL types possess identical attributes and differ only in epitope specificity. (a) Bubble chart representation depicting the distribution of the various CTL types in CTL 'fitness space' and their respective relative abundances for each of the five sample data sets. (b) List plot representation showing the relative immunogenicities and prevalences corresponding to the CTL types for each of the sample data sets. (c) The curves for each sample data set describing the relationship between the two measures of viral burden, the proviral load, v^*, and the active proportion of proviral cells, $y^*/(u^*+y^*)$, as parametric functions of the rate of spontaneous infected target cell activation, τ, representing the fitness of the virus. Each of the curves is obtained by numerically solving model (5.1) at equilibrium with τ as the bifurcation parameter, and plotting the relationship between v^* and $y^*/(u^*+y^*)$. In the above, for a given sample, all CTL types are ordered sequentially with respect to increasing immunogenicity. Parameter values and the corresponding relative abundances for each of the CTL types are shown for reference in Table 5.4, while the measures of viral detriment for selected rates of spontaneous proviral expression are shown in Table 5.5.

To make comparisons amongst the five different sets of sample data, we first construct visual representations describing various attributes of each CTL response, namely, the relative immunogenicity of an epitope, the effective expansion rate for its cognate CTL type, the efficiency of CTL-mediated lysis, and the relative abundance of that particular CTL type. Figure 5.5 shows 'colour bubble chart' and 'list plot' representations of the CTL repertoire for each of the five individuals as we described previously in Section 5.2.4 (see also Table 5.4 for the assigned parameter values used to create the graph). An alternative visualisation places all the CTL responses from the five different individuals in the same 'CTL fitness space' or along the same axis with respect to relative immunogenicity, as shown in Figures 5.6(a)–(b). Next, in Figure 5.6(c), we examine the impact of cell-mediated immunity on the virus population for each of the five sets of sample data by plotting the corresponding curves relating the two measures of viral burden, namely the proviral load and the activated proportion of proviral cells. We discover, as would be expected, that differing immune responses between individuals establish different levels of control on the virus population. Moreover, we notice that because of such heterogeneity between individuals with respect to immune-mediated control, the same fixed rate of spontaneous viral expression given by the parameter τ in the multiple epitope model (5.1) can differentially manifest itself in terms of viral burden, analogous with what we have seen previously within a single individual (we refer the reader to the open triangle $\triangle$, open square $\square$, and open circle $\bigcirc$ markers in Figure 5.6(c); see also Table 5.5 for comparisons of the manifestation of viral detriment given fixed rates of spontaneous proviral expression, i.e. small, intermediate and large values of the parameter τ, in each of the five sample data sets). For instance, we see that a specific rate of proviral expression (e.g. $\tau = 0.030$) is associated with an active proportion of infected cells and proviral load for sample data SD-1 which is higher than those for any of the other sets of sample data (compare the positions of the open squares $\square$ for each of the five sample data sets in Figure 5.6(c)).

The next question that one might ask is whether we can identify which characteristics constitute a 'good' cellular immune response, in other words, one that is effective in maintaining a low degree of viral detriment? Taking a more careful look at Figure 5.6(c), we immediately notice that sample data sets SD-1 and SD-2 display the highest degrees of viral detriment, whilst sample data sets SD-3 and SD-4 both exhibit the lowest degrees of viral burden and, moreover, are at similar levels (refer to Tables 5.4 and 5.5 for a comparison of the numerical values of the parameters and quantities that define the immune profile and viral burden for each sample data set). The control sample data set SD-C lies in between the two extremes at an intermediate level of viral burden. Considering that the CTL response is different in each of the above three sample data sets, we wonder why is it that SD-1 and

SD-2 are poor at controlling the virus compared to SD-3 and SD-4, and additionally, how do both SD-3 and SD-4 achieve a similiar degree of immune-mediated suppression despite having vastly distinct CTL repertoires? Referring to the CTL attributes for each of the sample data sets in Figure 5.5, we observe that the CTL response for sample data SD-1 consists principally of a single immunodominant CTL type 10 which has the highest rate of effective expansion and targets a highly immunogenic epitope (specifically, the third most immunogenic epitope 10), but has a very low rate of lysis (in fact, the second lowest CTL killing rate out of all the CTL types). Meanwhile, for sample data SD-2, which does a slightly better job of controlling the virus than SD-1, the cellular immune response is much more heterogeneous, with multiple CTL types existing at similar prevalences. However, along with low rates of CTL expansion across all CTL types, the distribution of relative immunogenicities for SD-2 means that the more efficient CTL types do not prevail at high enough abundances to effectively control the virus. Nevertheless, comparing sample data sets SD-1 and SD-2 suggests that a more balanced CTL population (SD-2) that can represent a wide range of efficiencies is better at reducing viral burden than a single large, inefficient CTL response (SD-1), as would be expected from intuition (compare the blue (SD-1) and violet (SD-2) curves in Figure 5.6(c)). This observation is further exemplified by comparison with the control sample data SD-C, whose CTL response exhibits a uniform distribution of effector cells in that all CTL types have equal immunogenicity and equal effective expansion rates, and hence equal abundances. As a result, the more efficient CTL types 5 and 1, which are poorly immunogenic for SD-1 and SD-2, are more immunogenic for SD-C and can therefore exert a greater impact on reducing the proviral load and activated proportion of proviral cells.

By contrast, the sample data sets SD-3 and SD-4 represent two different examples of immune profiles that are able to effectively suppress the viral infection (refer to Figure 5.5, third (SD-3) and fourth (SD-4) columns; also, Figure 5.6, yellow (SD-3) and green (SD-4) colours). On the one hand, the immune response from sample data set SD-3 shows co-dominance of three CTL types 8, 11, and 12 that exist at similar abundances, arising from a trade-off between effective expansion and relative immunogenicity. Whereas CTL types 11 and 12 correspond to targets that are highly immunogenic, they also expand at a slower rate than CTL type 8, which has a faster rate of effective expansion, but targeting a less immunogenic epitope. The balance between the two forces (i.e. immunogenicity versus effective expansion rate) creates the three similar-sized responses, two of which are likely to exert the greatest control over viral burden, namely, the less immunogenic but quickly expanding, highly efficient CTL type 8, and the more immunogenic but slower expanding, moderately efficient CTL type 11 (we note that the slowly expanding, inefficient CTL type

12 is unlikely to have a large impact on controlling the viral burden, despite its high degree of immunogenicity). Lastly, we make a comparison between these two sample data sets SD-3 and SD-4, both of which are equally effective at controlling viral burden. We observe that the CTL repertoire in sample data set SD-4 appears more evenly distributed than SD-3, and much like the control data set SD-C, the immunogenicities for the various CTL responses are similar to each other and so the prevalence of each CTL type relative to the others depends more on its individual rate of effective expansion. In this case, many of the CTL types exist at similar prevalences, however, since the more efficient responses, such as CTL type 1 and CTL 11, also expand at higher rates, their presence allows them to have a greater effect on the virus population.

As a side remark, we notice that the CTL distribution of the immune response for sample data SD-4, which is effective at suppressing viral burden, appears similar to that of SD-2, whose control of the virus is poor (see Figure 5.5, second (SD-2) and fourth (SD-4) columns; also, Figure 5.6, violet (SD-2) and green (SD-4) colours). What is it, then, that distinguishes these two sets of sample data? Although both sets exhibit a fairly good mix of all CTL types, the key difference is that for the better controller of viral burden (SD-4), the rates of expansion of efficient (e.g. CTL type 1) or at least moderately efficient (e.g. CTL type 8) CTL responses are faster than those of the CTLs from the poor controller of viral detriment (SD-2). This contrast between sample data SD-2 and SD-4, in tandem with the comparison between sample data SD-3 and SD-4, fortifies the notion that only a handful of CTLs (and not the entire CTL repertoire) are the main driving force behind effective viral control.

Lastly, we re-iterate the observation that the cellular immune response for the control sample data (SD-C), which is comprised of a uniform mix of CTLs that all have identical effective expansion rates and are equally immunogenic (and hence exist at equal relative abundances according to Equation (5.5)), is better at controlling viral detriment than sample data SD-1 and SD-2, both of which exhibit much more heterogeneity in their respective CTL repertoires than SD-C, both in terms of individual effective CTL expansion rates as well as the distribution of relative immunogenicities for each CTL type. A plausible reason for the inefficacy in sample data SD-1 and SD-2, as we have mentioned above, is that efficient CTL types do not exist at high enough prevalences to have a significant effect on viral burden. Thus, it would make sense that increasing the abundance of efficient CTLs targeting poorly immunogenic epitopes could result in a reduction in the proviral load and activated proportion of proviral cells. To test this hypothesis, we consider the five sets of sample data SD-1, SD-2, SD-3, SD-4 and SD-C, and in the context of our multiple epitope model (5.1), we adjust the immunogenicity of all epitopes to be equal; that is,

in each of the sample data sets, $\varepsilon_k = \varepsilon$ for all $k = 1, \ldots, 12$. The results are plotted in Figure 5.7 and 5.8 (refer to Table 5.4 and Table 5.5 for the various parameter values used, the relative abundances of each CTL type, and the measures of viral burden corresponding to selected rates τ of spontaneous proviral expression). What we notice immediately is that the magnitude of the proviral loads for the worst controller of viral burden, SD-1, is significantly lower for the case when all CTL responses have equal immunogenicity than in the case with the original immunogenicity profile (compare Figures 5.6(c) and 5.8(c), blue lines). A closer look at the numerical values for the proviral loads in Table 5.5 reveals that equal immunogenicity amongst all epitopes leads to a lower proviral load for a given rate of spontaneous proviral expression, although there is little change in the second measure of viral burden, the active proportion of proviral cells. The reason for this is not that a uniform immunogenicity profile is necessarily optimal in reducing the proviral load in all cases, but rather because levelling out the immunogenicities of all CTL types means that efficient responses targeting poorly immunogenic epitopes can exist at higher abundances and therefore play a greater role in eliminating virus-infected cells.

The results discussed here are consistent with our previous examination of different immununogenicity profiles in determining the structure and impact of anti-HTLV-I cellular immunity. In particular, the findings from our multiple epitope model generalises and supports the conclusions from the baseline and Tax/HBZ models that an effective immune response is one that selects for CTL types that are efficient at killing infected cells, expand rapidly, and whose corresponding epitope targets are sufficiently immunogenic to elicit a sizeable response in order to cause a reduction in the magnitude of the proviral population.

Sample Data	CTL Property	CTL Type											
		1	2	3	4	5	6	7	8	9	10	11	12
	CTL Effic.	0.130	0.113	0.089	0.028	0.142	0.069	0.061	0.113	0.123	0.028	0.095	0.027
SD-1	Eff. Exp.	3.643	3.373	0.595	1.937	0.254	2.741	1.668	0.240	2.635	5.897	0.975	1.114
	Immunog.	0.00656	0.00885	0.0109	0.0184	0.0492	0.0684	0.0830	0.111	0.115	0.144	0.191	0.193
	Abun.	0.0119	0.0148	0.00323	0.0177	0.00621	0.0930	0.0687	0.0132	0.150	0.422	0.0923	0.107
	Abun. (IP-E)	0.145	0.135	0.0237	0.0773	0.0101	0.109	0.0665	0.00955	0.105	0.235	0.0389	0.0444
SD-2	Eff. Exp.	1.732	2.339	0.715	2.741	0.499	3.105	2.797	2.790	1.232	3.461	3.091	1.677
	Immunog.	0.0143	0.0193	0.0517	0.0682	0.0724	0.0748	0.0911	0.0941	0.101	0.105	0.140	0.168
	Abun.	0.0108	0.0198	0.0162	0.0819	0.0158	0.102	0.112	0.115	0.0546	0.160	0.189	0.124
	Abun. (IP-E)	0.0662	0.0893	0.0273	0.105	0.0191	0.119	0.107	0.107	0.0471	0.132	0.118	0.0640
SD-3	Eff. Exp.	2.379	5.254	2.408	1.308	0.255	2.134	4.344	10.421	1.034	1.106	4.310	4.894
	Immunog.	0.000502	0.0281	0.0282	0.0415	0.0441	0.0509	0.0552	0.0764	0.0985	0.113	0.225	0.238
	Abun.	0.000315	0.0389	0.0179	0.0143	0.00297	0.0287	0.0632	0.210	0.0269	0.0331	0.256	0.308
	Abun. (IP-E)	0.0597	0.132	0.0604	0.0328	0.00641	0.0535	0.109	0.262	0.0259	0.0278	0.108	0.123
SD-4	Eff. Exp.	6.005	4.196	3.990	0.645	1.621	1.372	3.044	2.596	2.586	5.768	6.626	3.984
	Immunog.	0.0499	0.0520	0.0627	0.0730	0.0805	0.0858	0.0903	0.0909	0.0993	0.103	0.105	0.107
	Abun.	0.0845	0.0614	0.0705	0.0133	0.0368	0.0332	0.0775	0.0665	0.0724	0.168	0.196	0.121
	Abun. (IP-E)	0.142	0.0989	0.0940	0.0152	0.0382	0.0323	0.0717	0.0612	0.0609	0.136	0.156	0.0939
SD-C	Eff. Exp.	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4
	Immunog.	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12
	Abun.	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12
	Abun. (IP-E)	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12	1/12

Table 5.4: CTL properties for each of the $p = 12$ CTL types in five sets of sample data simulated from the multiple epitope model (5.1) corresponding to Figures 5.5, 5.6, 5.7 and 5.8. Each CTL is defined by four attributes: (i) relative immunogenicity, (ii) effective expansion rate, (iii) efficiency of CTL-mediated lysis, and (iv) relative abundance. For each CTL property, the three largest values are highlighted in cyan (first), lime (second), and yellow (third) for emphasis. The values for each of the parameters have been selected from biologically realistic ranges as in Table 3.1.

Sample Data	Immunogenicity Profile	Proviral Load			Active Proportion		
		Small τ ($\tau = 0.005$)	Intermediate τ ($\tau = 0.03$)	Large τ ($\tau = 0.12$)	Small τ ($\tau = 0.005$)	Intermediate τ ($\tau = 0.03$)	Large τ ($\tau = 0.12$)
SD-1	IP-Original	0.0209	0.0468	0.0327	0.0420	0.0797	0.191
	IP-E	0.0158	0.0364	0.0255	0.0415	0.0767	0.183
SD-2	IP-Original	0.0161	0.0370	0.0259	0.0415	0.0769	0.183
	IP-E	0.0154	0.0355	0.0248	0.0414	0.0764	0.182
SD-3	IP-Original	0.00943	0.0226	0.0159	0.0408	0.0728	0.171
	IP-E	0.00905	0.0218	0.0153	0.0408	0.0725	0.170
SD-4	IP-Original	0.00926	0.0222	0.0157	0.0408	0.0726	0.171
	IP-E	0.00864	0.0208	0.0147	0.0407	0.0723	0.170
SD-C	IP-Original	0.0127	0.0297	0.0209	0.0411	0.0748	0.177
	IP-E	0.0127	0.0297	0.0209	0.0411	0.0748	0.177

Table 5.5: Magnitude of the proviral load and the active proportion of proviral cells at equilbrium for selected levels of viral fitness (in terms of the extent of spontaneous proviral activation τ, which determines the basic reproduction number R_0 of the infection) for each of the $p = 12$ CTL types in five sets of sample data simulated from the multiple epitope model (5.1) corresponding to Figures 5.6(c) and 5.8(c).

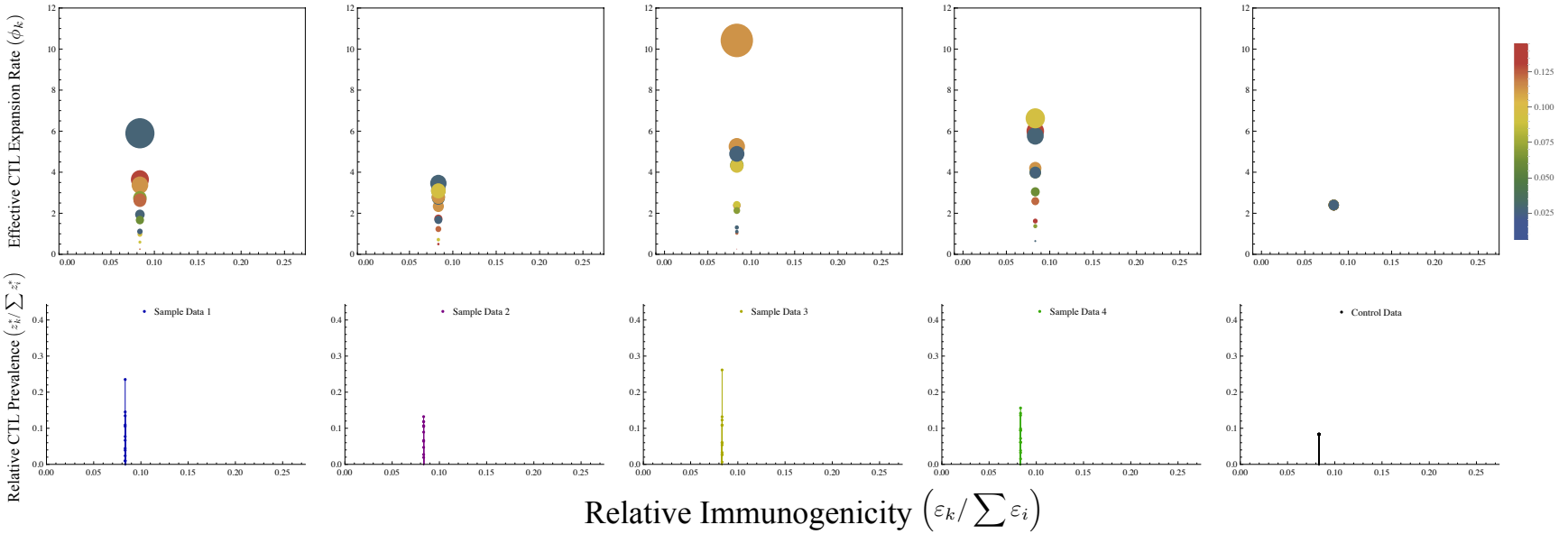
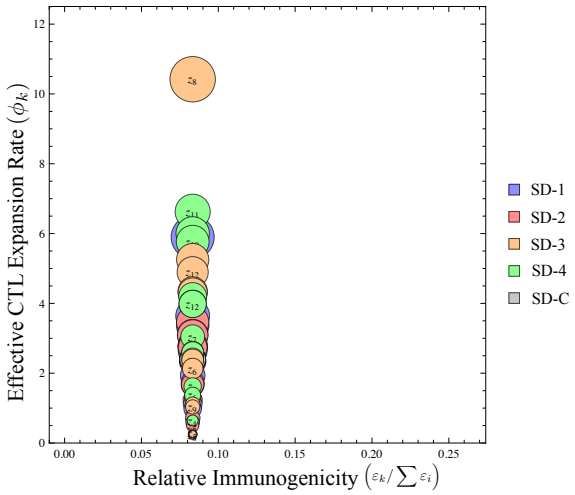
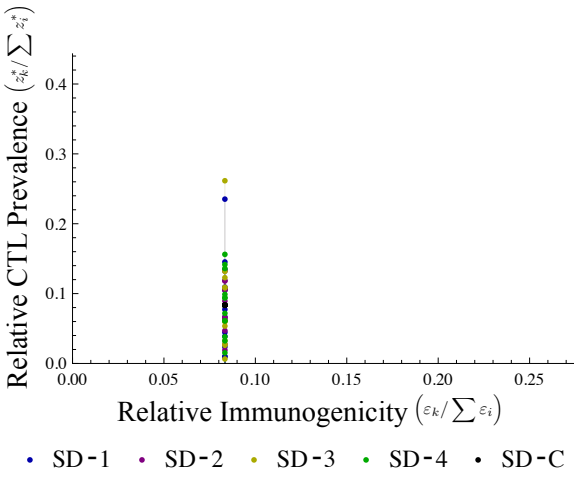


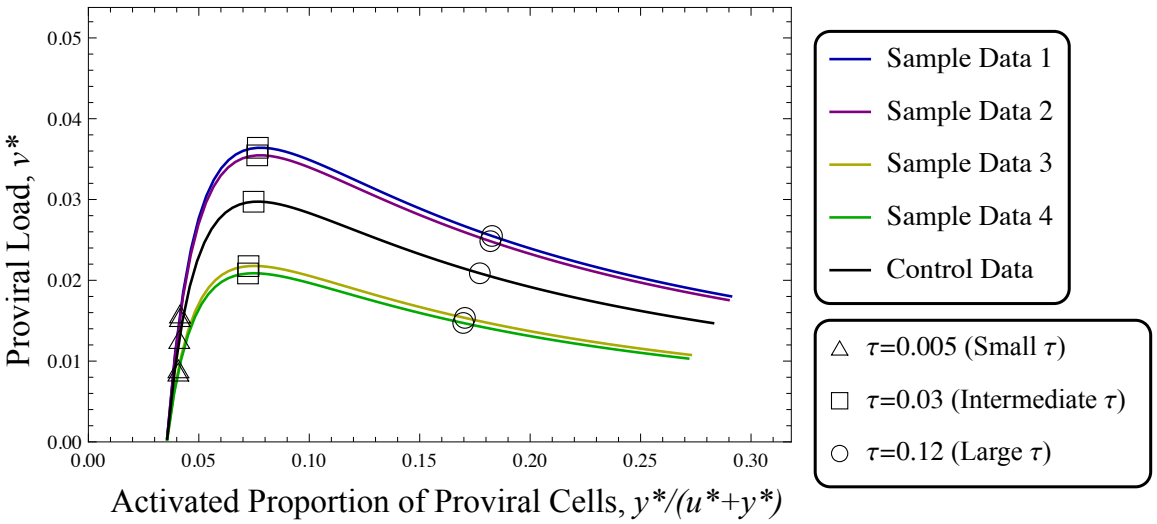
Figure 5.7: CTL properties and immunogenicity profiles for the immune responses of the five individuals described in Figures 5.5 and 5.6, except with uniform immunogenicity profiles. The graphs in each column show visual representations as described in Section 5.2.4 and correspond to the same set of sample individuals considered earlier (SD-1, SD-2, SD-3, SD-4, and SD-C). Here, the relative abundance of each CTL type depends only on its respective effective expansion rate. Top Row: The 'colour bubble chart' representations for each sample data set illustrate four CTL attributes: (i) relative immunogenicity (horizontal axis), (ii) effective expansion rate (vertical axis), (iii) efficiency of CTL-mediated lysis (colour of bubble), and (iv) relative abundance (size of bubble). Bottom Row: The corresponding 'list plot' representations for each sample data set highlight the relationship between the immunogenicity profile and the relative abundances of each CTL type. Values for the parameters and the corresponding relative abundances of effector T-cells defining the CTL profiles for each of the individuals are shown for reference in Table 5.4).



(a) Bubble chart indicating the position of each CTL type in CTL fitness space.



(b) Immunogenicity profile of each CTL type.



(c) Proviral load and active proportion as parametric functions of τ, the rate of spontaneous proviral activation and viral protein expression.

Figure 5.8: CTL dominance hierarchy and its effect on the virus population according to the multiple epitope model (5.1) across the same five sample individuals considered in Figures 5.5 and 5.6, but with uniform immunogenicity profiles. Each of the colours represents a particular sample data set as follows (unless otherwise indicated): SD-1 (blue), SD-2 (violet), SD-3 (yellow), SD-4 (green), SD-C (grey/black). In this case, the abundance of each CTL response is strictly dependent on its respective rate of effective expansion. We previously observed that for a general immunogenicity profile, CTL responses targeting poorly immunogenic epitopes were almost non-existent; however, with equal immunogenicity, these responses, which can sometimes be highly efficient at eliminating virus-infected target cells, can prevail at sufficiently large abundances to affect the size of the proviral load (compare and contrast Figures 5.6 and 5.8; also refer to the numerical values in Tables 5.4 and 5.5). This is apparent especially in sample data set SD-1 (c.f. Figures 5.6(c) and 5.8(c), blue lines). Parameter values and the resulting relative abundances for each of the CTL types are shown for reference in Table 5.4, while the measures of viral detriment for selected rates of spontaneous proviral expression are shown in Table 5.5.

5.4 Comparison of Model Outcomes with Data

In the previous sections we have developed a theoretical framework to explore the role that a diverse CTL response targeting multiple epitopes has in establishing a CTL immunodominance hierarchy, and how it can lead to differential control of the infection with respect to two measures of viral burden, the activated proportion of proviral cells and the size of the proviral load. Here, we apply our framework to published experimental data from the literature, which in several cases allows us to estimate parameter values related to CTL and virus properties. The ideas presented in this section are preliminary and require further detailed investigation. Nevertheless, the purpose of this exposition is to provide a proof of concept that our modelling framework can be used in conjunction with experimental data to help estimate parameter values that have not been, or cannot be, measured and, moreover, to draw conclusions that could have implications for future experimental designs and theoretical hypotheses.

5.4.1 Experimental Data From Published Literature

There is a considerable amount of published experimental data on HTLV-I. An HTLV-I seropositive individual is often uniquely identified by a series of anonymised two or three letter codes which are prefixed by 'T' for HAM/TSP patients and 'H' for healthy carriers. These identifying codes appear in various sources, with the same code representing the same infected individual (although blood samples may, in some cases, have been taken longitudinally — in other words, from a single person but at different points in time) (Asquith et al., 2005a,b; Daenke et al., 1996; Goon et al., 2004a; Parker et al., 1992; Niewiesk et al., 1994). We focus on those data that are reported as measurements of viral or CTL abundance isolated from fresh blood samples of HTLV-I seropositive individuals which are directly related to quantities that can be captured using our mathematical modelling framework. In the present context, we are interested in examining the interaction between the diverse HTLV-I-specific CTL response and our two measures of viral detriment, and we have focussed on data from three relevant sources, namely Goon et al. (2004a) and Asquith et al. (2005a,b).

Goon et al. (2004a) have performed a systematic investigation of the diversity in anti-HTLV-I CTL responses, creating a detailed immunodominance hierarchy for each infected individual in a cohort of 17 HTLV-I seropositives, 10 patients with the inflammatory disease HAM/TSP and 7 asymptomatic carriers, by measuring the frequencies of epitope-specific CTLs across almost the full range of HTLV-I antigens comprising eight viral proteins in total: Tax, Env, Rex, Tof, Rof, Gag, Pol and Pro. Table 5.6 reproduces the relevant data from Goon et al. (2004a) detailing the experimentally measured values for individual

epitope-specific $CD8^+$ T-cell counts corresponding to the various peptide targets for each HTLV-I-infected subject in the cohort, as well as their respective proviral loads (the original table is provided in Appendix C.2 for reference). We note that for each of the subjects in the study, there are two columns for the Tax protein, Tax13 and Tax20; however, both of these are meant to quantify the same epitope-specific CTL response, corresponding to those CTLs that target peptide fragments covering all of the Tax protein, with the only difference being in how the lengths of the peptide chains are cut. Specifically, Tax13 consists of peptides from the Tax protein that have 13mer[9] lengths overlapping by 9 amino acids, whereas Tax20 comprises peptides from Tax that have 20mer lengths overlapping by 14 amino acids. Hence, the measured Tax-specific CTL frequency for the Tax13 and Tax20 columns as reported in Table 5.6 describe the same response, and the reason for the slight differences in values between the two columns is due to varying degrees of recognition of the different-sized peptides by Tax-specific $CD8^+$ CTLs, which can come about, for example, if there were to exist an epitope on a given 13mer chain that can be recognised with high affinity by anti-Tax CTLs, but this same epitope is 'broken' when cutting the protein into 20mer length peptides. Although the measurements for Tax-specific $CD8^+$ CTL abundance tend to be slightly lower using 20mer peptide fragments instead of 13mer peptide chains, Goon et al. (2004a) demonstrated that the difference is not significant. In order to better represent the immunodominance of the Tax-specific CTL response, for the remainder of our analysis, we will consider the CTL abundances corresponding to Tax13 rather than Tax20. All other epitope-specific CTL responses have been measured after stimulation with 20mer length peptide fragments.

Meanwhile, Asquith et al. (2005b) have quantified two separate measures of viral burden, namely the proviral load with respect to the number of Tax copies per 100 PBMCs as well as the extent of viral expression in terms of the percentage of Tax^+ $CD4^+$ helper T-cells in the target cell pool. This study, along with an additional experimental study by the same research group, set out to measure and report the total rate of CTL-mediated lysis by HTLV-I-specific effector cells as defined by the proportion of Tax^+ $CD4^+$ helper T-cells that are cleared per $CD8^+$ cell per day (Asquith et al., 2005a,b), thereby characterising the overall efficiency of the HTLV-I-specific CTL response. We combine and list the relevant information from Asquith et al. (2005a,b) reporting the experimental measurements of viral burden and the overall HTLV-I-specific CTL lysis rates in Table 5.7 (the original full tables are provided in Appendix C.2 for reference).

[9]The suffix '-mer' refers to the length of the peptide chain with respect to the number of amino acids present. A 13mer length peptide is composed of 13 amino acids, whereas a 20mer length peptide consists of 20 amino acids.

Clinical Status	Subject/ Patient	PVL (% PBMCs)	Peptide									
			Tax13	Tax20	Env	Rex	Tof	Rof	Gag	Pol	Pro	Total
HAM/TSP	TAY	14.00	8480	ND	5590	0	1540	0	3000	7880	0	26500
	TBA	10.90	631	496	2000	0	996	673	0	902	0	5220
	TAQ	1.50	584	ND	0	0	0	0	0	0	0	584
	TAU	2.70	534	ND	300	152	243	265	117	2350	0	3961
	TBI	32.00	3390	ND	0	0	0	0	0	1230	0	4620
	TW	23.00	6620	3450	980	0	0	0	392	0	0	7990
	TBK	35.20	4900	2390	2060	0	0	0	416	537	416	8330
	TAN	6.20	29000	34000	3300	0	0	0	0	0	0	37200
	TAE	3.10	7850	8220	337	0	214	241	228	296	0	9170
	TAL	26.20	4460	3160	617	0	0	0	225	0	0	5300
AC	HBD	0.01	8130	6430	0	0	0	0	394	947	0	9470
	HBE	11.50	8960	6160	2170	0	0	563	2290	883	374	15000
	HBF	6.50	857	690	0	0	0	290	273	623	0	2040
	HT	1.20	15900	5300	1680	0	0	0	0	4940	0	22520
	HAY	10.60	4110	0	0	0	0	0	0	0	1570	5680
	HX	0.01	4670	2960	772	0	2320	822	1810	4530	0	14924
	HAE	0.01	0	0	0	0	0	0	0	5690	0	5690

Table 5.6: Experimental data from Goon et al. (2004a). The table shows measured $CD8^+$ CTL frequencies for an almost full range of HTLV-I epitopes (eight in total), thus creating a CTL immunodominance hierarchy across 17 HTLV-I seropositive individuals, including 10 HAM/TSP patients (prefixed with 'T') and 7 healthy carriers (prefixed with 'H'). We will use these measured values for parameter estimation using our multiple epitope model framework in Section 5.4.3. This table is adapted from the one found in Goon et al. (2004a) (Table 1), which is included for reference in Appendix C.2. All non-Tax peptides are 20mers in length. '0': Below the limit of detection; ND: Not done; Tax13: Tax 13mer peptides (overlapping by 9 amino acids); Tax20: Tax 20mer peptides (overlapping by 14 amino acids).

Clinical Status	Subject/ Patient	Total Effective Rate of CTL Lysis (per $CD8^+$ cell per day)	% Tax Expression (Tax+ CD4+/CD4+)	Proviral Load (% PBMCs)
HAM/TSP	TAC	0.091	11.1	12.1
	TAY	0.298	4.4	2.2
	TBA	0.050	6.1	3.5
	TAT	0.058	9.6	4.2
	TBG	0.007	21.5	16.4
	TW	0.024	12.4	10.3
	TAQ	0.083	2.7	1.0
	TAU	0.049	9.8	5.4
	TBI	0.003	8.1	12.3
	TAL	0.053	–	ND
	TAA	0.048	–	ND
	TBJ	0.039	–	ND
	TBO	0.172	–	ND
	TAZ	0.020	–	ND
AC	HS	0.001	3.0	5.8
	HT	0.062	3.1	1.0
	HY	0.065	0.6	0.4
	HAY	-0.007	7.5	10.6
	HBD	0.220	0.1	0.001
	HBF	0.029	2.9	2.8
	HBH	0.020	2.7	3.98
	HCB	0.051	–	ND
	HBE	0.427	–	ND

Table 5.7: Combined experimental data from Asquith et al. (2005a,b). The table shows experimentally obtained values for the total effective, or overall, rate of HTLV-I-specific CTL lysis, the percentage of $CD4^+$ T-cells expressing Tax, and the proviral load in terms of the proportion of Tax copies per 100 PMBCs. The shaded rows indicate those HTLV-I seropositive individuals who also appear in the experimental data in Goon et al. (2004a). For such individuals, their respective immunodominance hierarchies are known (see Table 5.6). The original tables from which this table is adapted are included for reference in Appendix C.2. ND: Not done. '–': Data not available.

In the following subsection we will link these data to corresponding quantities in our multiple epitope modelling framework.

5.4.2 Linking Experimental Measurements to Model Quantities

The measured data values from Goon et al. (2004a) and Asquith et al. (2005a,b) correspond directly to quantities that appear in our multiple epitope model (5.1). We would therefore like to incorporate the data into our mathematical framework and use the model to explore aspects of realistic HTLV-I infection in the set of sample patients.

With respect to the experimental data of Goon et al. (2004a), the measurements for the abundances of individual epitope-specific effector cells correspond precisely to the quantity z^*_{epitope} over the set of all epitopes in our multiple epitope model (5.1). Because Goon et al. (2004a) considered eight distinct epitopes in total, we take $p = 8$ in the model to coincide with the data. Meanwhile, the experimental data of Asquith et al. (2005a,b) captures three quantities that are analogous to the expressions appearing in our model. First, the measurement for overall CTL efficiency represents the average rate of lysis across all HTLV-I epitopes, which can be represented by the single parameter γ[10]. Next, since the measurements of the percentage of Tax expression in the $CD4^+$ T-cell pool have been taken during the chronic phase of infection long after establishment of the virus (often years or even decades following initial infection), we assume that they have reached a steady state. These measurements correspond not exactly to the activated proportion of proviral cells $y^*/(u^* + y^*)$ at equilibrium that we have derived analytically in our multiple epitope model (5.1), but to another related quantity describing the activated infected fraction of all $CD4^+$ helper T-cells (including healthy $CD4^+$ T-cells), given by the expression $y^*/(x^* + u^* + y^*)$ in our model. Lastly, in the experimental data, the proviral load has been expressed as a percentage of PBMCs that are infected, whereas in our model, we only consider the pool of $CD4^+$ helper T-cells. Nevertheless, because the virus primarily resides in the subset of $CD4^+$ helper T-cells, the proviral load defined with respect to the infected proportion of PBMCs is directly proportional to the proviral load according to our characterisation (given by the expression v^* in our mathematical model), scaled only by a constant factor depending on the state of the individual[11]. As our model describes only the behaviour of the $CD4^+$ helper T-cell pool, we apply a constant scaling factor of 0.3 $CD4^+$/PBMC to the data[12]. To obtain the adjusted values for the proviral load in terms of Tax copies per

[10]Note, however, that the individual lysis rates corresponding to a particular epitope-specific CTL type has not been determined.

[11]$CD4^+$ T-cells make up between 25–60% of the lymphocyte population (Bofill et al., 1992), which themselves comprise 20–50% of peripheral blood leukocytes (i.e. PBMCs) (Parham, 2005).

[12]This constant scaling factor has been obtained by taking the upper limit for the $CD4^+$ T-cell proportion in the pool of PBMCs. Specifically, we assume that $CD4^+$ T-cells comprise 60% of the lymphocyte population

100 $CD4^+$, we take the measured value for the proviral load in terms of Tax copies per 100 PBMCs, and divide this value by 0.3 $CD4^+$/PBMC (refer to Figure 5.9(c) for a bar chart showing the adjusted proviral load measurements from Asquith et al. (2005a,b), which we will use for our parameter estimation procedure in the next subsection).

Both Goon et al. (2004a) and Asquith et al. (2005a,b) report values for the proviral load for the majority of the infected individuals, and there exists some overlap in the patients; that is, a few of the samples originating from the same patient were examined in both studies. However, we observe that although the experimental procedure for determining the proviral load appears to be the same for the two studies (Asquith et al., 2005a,b; Goon et al., 2003, 2004a), in several instances there is a discrepancy in the measurements for the proviral loads of the same individual between the studies (compare the corresponding entries for the PVL in Table 5.6 and Table 5.7). The reason for this discrepancy is likely due to the fact that the two experiments were carried out at different points in time, even though the set-point proviral load within an individual is known to remain stable over time (Asquith and Bangham, 2008). Indeed, since the proviral load has been expressed in terms of infected PBMCs, the presence or absence of other PBMCs in serum samples can result in a wide variation in the measurement of the proviral load at a given time. These other PBMCs include, for instance, monocytes, macrophages, dendritic cells, or other lymphocyte subsets which are less likely to carry HTLV-I provirus such as B-cells, $CD8^+$ T-cells, and NK cells, and such variability in the composition of PBMCs, including the subset of $CD4^+$ T-cells, has been shown to occur even diurnally (Bofill et al., 1992).

Because the experimental design in Asquith et al. (2005b) was purposefully aimed at quantifying the virus-host interactions, and since the measurements of proviral load and the Tax^+ proportion of $CD4^+$ T-cells express two measures of viral burden that correspond directly to the two theoretically derived quantities from our multiple epitope modelling approach (and not just a single measurement of the proviral load), in order to utilise as much experimental information as possible, we shall use the proviral load measurements alongside the measurements of Tax expression reported by Asquith et al. (2005b) as in Table 5.7 when we incorporate the data values into our mathematical modelling framework in the following section.

The particular cohort of individuals that we will consider are the 17 HTLV-I seropositive subjects whose immunodominance hierarchy has been established by Goon et al. (2004a). Namely, 10 HAM/TSP patients identified by TAY, TBA, TAQ, TAU, TBI, TW, TBK, TAN, TAE and TAL, and 7 ACs identified by HBD, HBE, HBF, HT, HAY, HX and HAE.

and, moreover, that lymphocytes comprise 50% of all PBMCs. This implies that $CD4^+$ T-cells make up 30% of all PBMCs, i.e. 60 $CD4^+$/100 Lympocyte $\times$ 50 Lympocyte/100 PBMC = 30 $CD4^+$/100 PBMC, or 0.3 $CD4^+$/PBMC.

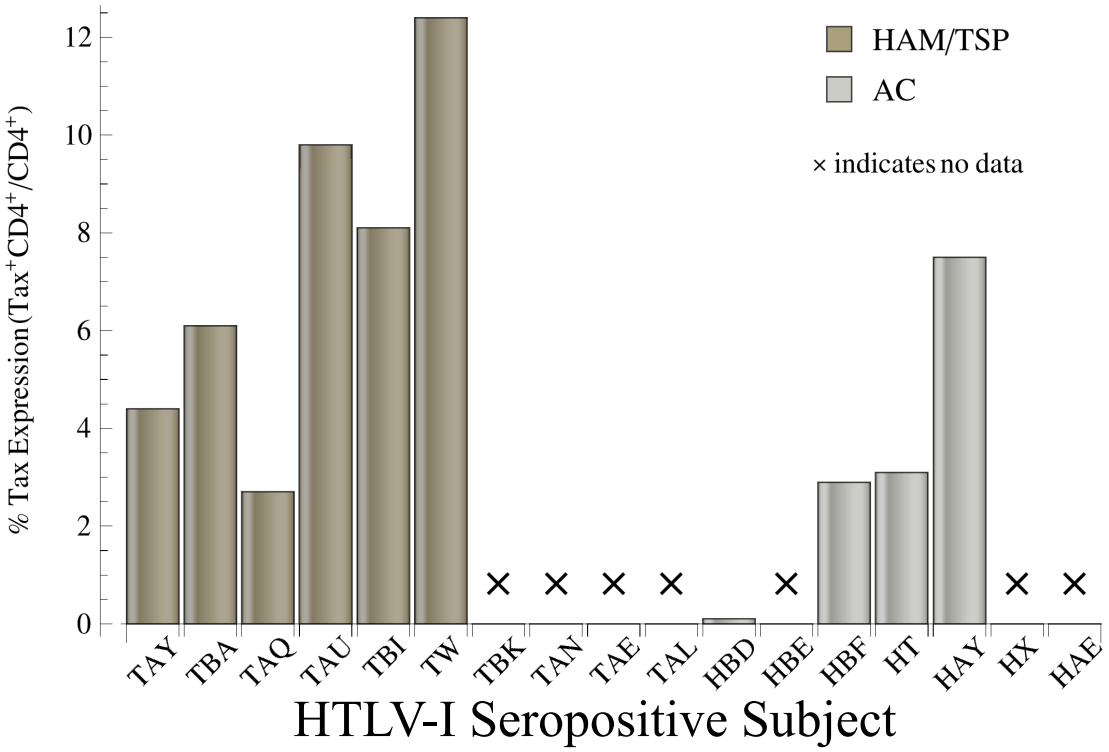
For a handful of these patients, the viral burden, both in terms of the proviral load as well as the extent of Tax expression in the $CD4^+$ cell pool, has been quantified experimentally and listed in Table 5.7 (shaded rows). Figure 5.9 shows a bar chart representation of the two measures of viral burden from Asquith et al. (2005b) for each of the HTLV-I-infected individuals considered by Goon et al. (2004a). In the subset of HAM/TSP patients, data were available for 6/10 subjects, and in the subset of ACs, data were available for 4/7 subjects. Figure 5.9(a) shows the percentages of Tax expression, whereas Figures 5.9(b)–(c) show the proviral load as reported by Asquith et al. (2005b). The proviral load is graphed both in terms of the infected percentage of PBMCs (Figure 5.9(b)) as well as with respect to the infected percentage of $CD4^+$ helper T-cells, adjusted by our previously described choice of constant scaling factor of 0.3 $CD4^+$/PBMC (Figure 5.9(c)). Where data for the extent of viral burden were not available, we have indicated this with an '$\times$' symbol. For bar chart representations of the above data separating the cohort by clinical status (i.e. HAM/TSP patients and ACs), see Appendix Figure C.4.

In the following section, we collate the relevant experimental data from the three sources mentioned here (summarised in Tables 5.6 and 5.7) and, after integrating them into our mathematical framework, we use the multiple epitope model (5.1) to derive estimates for two key parameters, namely, the average rate of effective CTL expansion (represented by ϕ in the model), and the rate of spontaneous infected target cell activation accompanied by the expression of viral proteins (represented by τ in the model), both quantities for which experimentally measured data are not readily available. The estimation of these parameters is an important step in understanding the role of a diverse immune response to HTLV-I infection.

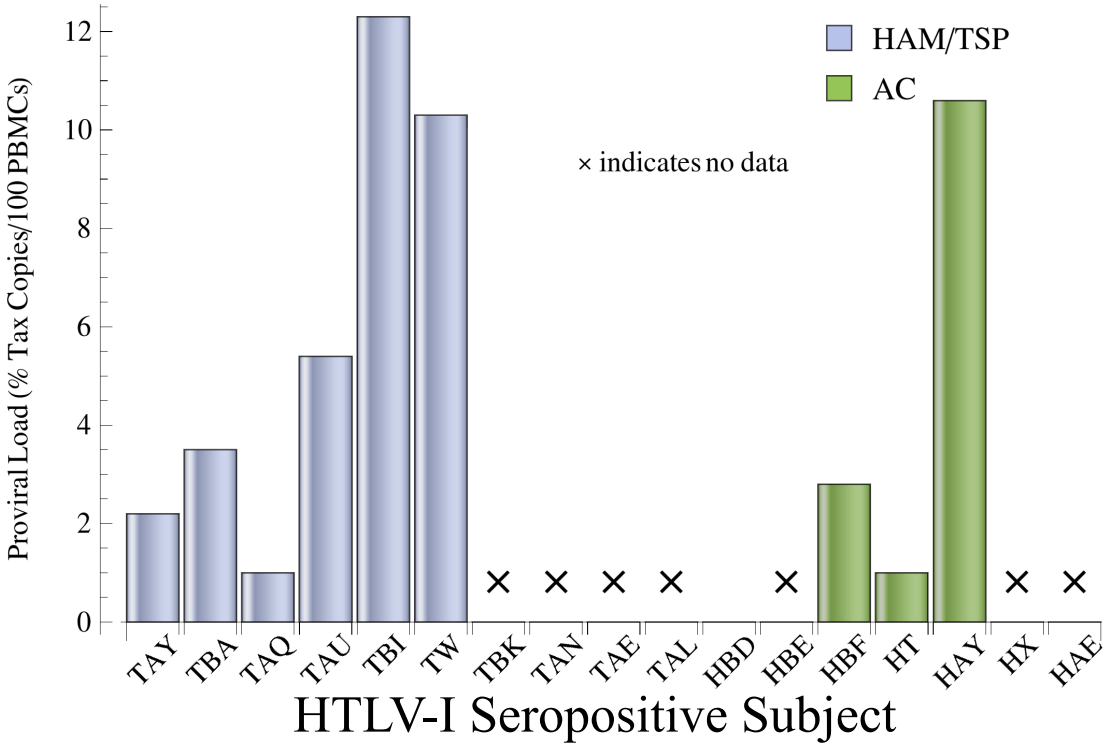
5.4.3 Parameter Estimation Using the Multiple Epitope Model

In the general multiple epitope model (5.1) with p epitopes, we have identified four key parameters (or parameter groupings) which characterise the host-virus interaction within different HTLV-I-infected subjects[13]. Specifically, viral fitness in terms of the rate of spontaneous proviral expression (τ), the immunogenicity of each of the epitopes (ε_k for $k = 1, \ldots, p$), CTL fitness in terms of the effective expansion rates for each CTL type (ϕ_k for $k = 1, \ldots, p$), and CTL efficiency in terms of the respective cell-mediated lysis rates for each CTL type (γ_k for $k = 1, \ldots, p$). Meanwhile, the experimental data set which we are considering yields measurements for the CTL immunodominance hierarchy consisting of eight different epitope-specific CTL types, the lysis rates for the Tax-specific CTL response

[13]Recall that following our investigations of the baseline model (3.1), we identified three of these critical parameters defining the host-virus interface to be spontaneous viral expression (i.e. τ), effective CTL expansion (i.e. ϕ), and CTL efficiency (i.e. γ).

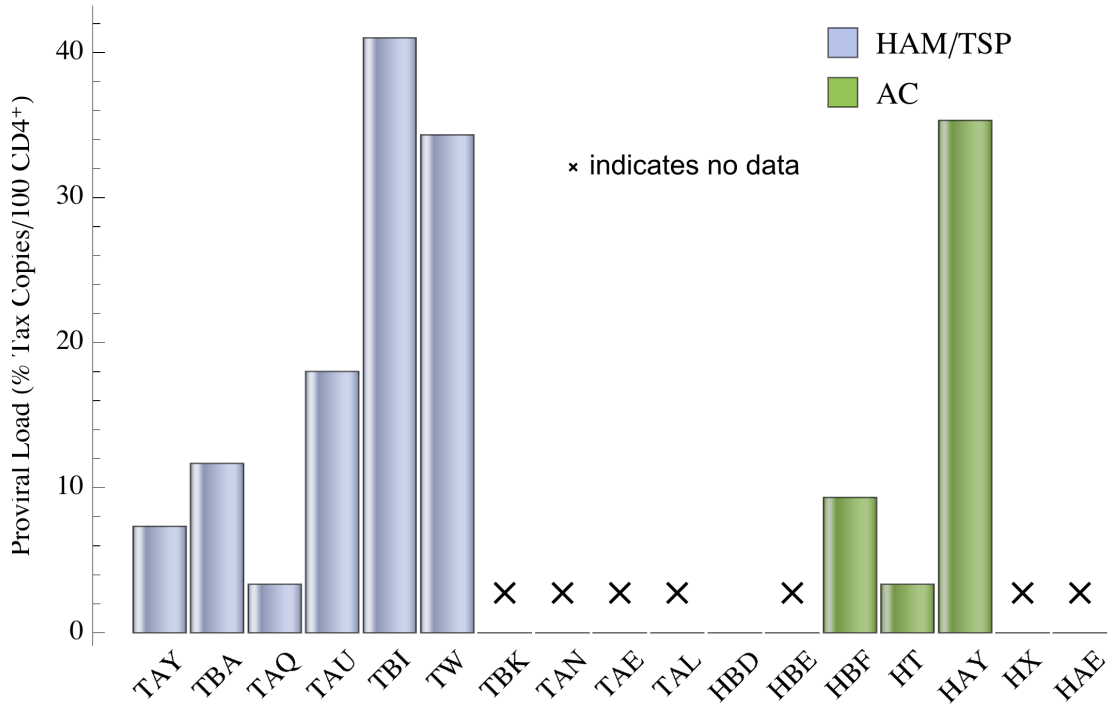


(a) Measured Tax^+ proportion of $CD4^+$ T-cells.



(b) Measured PVL in terms of percentage of Tax copies per 100 PBMCs.

Figure 5.9: **(Caption on following page.)**



(c) Adjusted PVL in terms of percentage of Tax copies per 100 $CD4^+$ T-cells.

Figure 5.9: Bar chart of the experimental data from Asquith et al. (2005a,b) describing the viral burden for a cohort of HTLV-I seropositive individuals. Only the values for patients appearing in Goon et al. (2004a) are considered (refer to Table 5.6 and Table 5.7, shaded rows). ×: indicates no data available for the particular individual. For bar chart representations of the above data separating the cohort by clinical status (i.e. HAM/TSP patients and ACs), see Appendix Figure C.4. Note: To obtain the adjusted PVL in (c), we considered a constant scaling factor of 0.3 $CD4^+$/PBMC and applied it to the measured PVL data shown in (b).

(though not for the CTL responses targeting non-Tax epitopes), and two measures of viral burden (i.e. the PVL and the activated proportion of proviral cells) for each infected subject in the cohort (refer to Tables 5.6 and 5.7).

In its current formulation, the general multiple epitope model (5.1) is over-parameterised for the amount of data that we have (e.g. for each infected subject, we consider the existence of eight distinct CTL types, but the experimental data only reports the total overall efficiency of the HTLV-I-specific CTL response). In order to incorporate as much of the experimental data as possible, we will have to introduce several justifiable assumptions to the mathematical model, which we will explain as we proceed through this section. In the cases where patient data are not available, we have chosen the corresponding parameter values from Table 3.1.

Our aim is to demonstrate a proof of concept that using our multiple epitope model (5.1),

we can utilise measured quantities to derive estimates for three of the four critical parameters (or parameter groupings) characterising the CTL properties of a given HTLV-I-infected subject for which no data are otherwise available, namely: (i) the immunogenicity profile, in terms of the relative immunogenicities of each epitope, denoted by $\varepsilon_k / \sum_{j=1}^{p} \varepsilon_j$ for $k = 1, \ldots, p$ (here $p = 8$), (ii) the rate of spontaneous viral expression, denoted by τ, and (iii) the overall rate of effective CTL expansion, which we denote by ϕ.

For the first, the experimental measurements in Table 5.6 detail the CTL immunodominance hierarchy; that is, the abundances of the various CTL effector types. Using these measurements in conjunction with the model, we can determine the immunogenicity profile that gives rise to the immunodominance hierarchy observed in the experimental data. For the second, spontaneous expression of the provirus which accompanies infected target cell activation is a measure of viral fitness and regulates the ability of the virus to propagate within the host in terms of the basic reproduction number for viral infection R_0, which in our mathematical model is defined by the expression in Equation (5.2). The rate of spontaneous viral expression is an intrinsic characteristic of the virus which determines viral burden with respect to the proviral load and the activated proportion of proviral cells as we have seen in the baseline model (3.1), and a high rate has been associated with the severity of HTLV-I-linked pathology (Asquith and Bangham, 2007, 2008). Lastly, for the third, measurements of virus-specific CTL growth focussed on quantifying the average rate of target cell expansion across all CTL types rather than individual measurements of the effective expansion rate for each particular epitope-specific CTL type (see Table 3.1). Yet, because epitope-specific CTLs differ primarily in specificity and effector functionality (i.e. rate of CTL-mediated elimination), as a simplification we assume that the effective expansion rates for all CTLs are identical, at a rate ϕ, so that $\phi_k = \phi$ for all $k = 1, \ldots, p$[14]. The determination of the above three parameters or parameter groupings, namely, $\varepsilon_k / \sum_{j=1}^{p} \varepsilon_j$, τ and ϕ, within our sample cohort can therefore help shed light on the virus-host interaction in chronic HTLV-I infection.

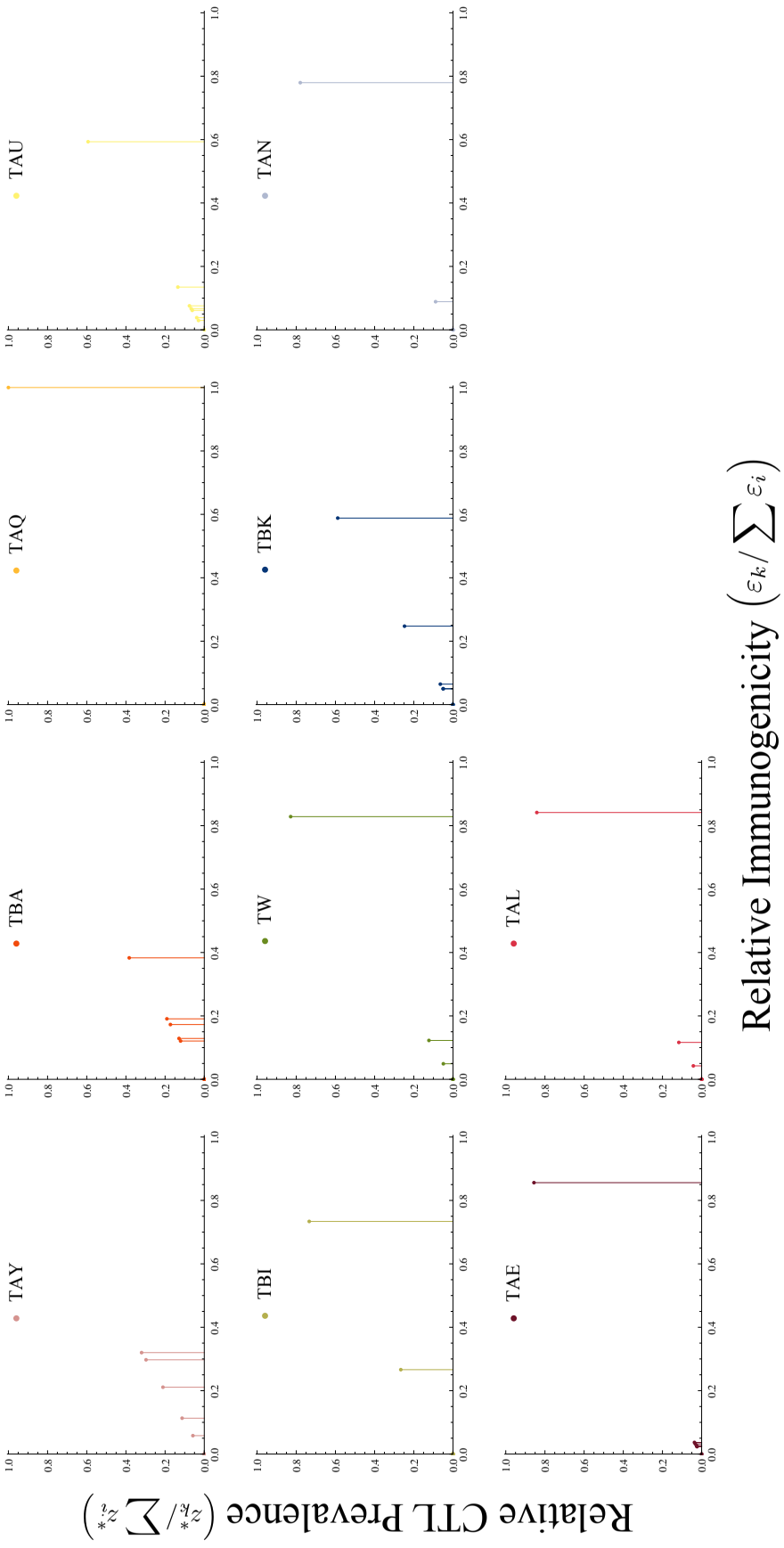
According to our multiple epitope model (5.1), the relative abundance for each CTL type k, where $k = 1, \ldots, p$ and p is the total number of epitope targets, is dependent on the weighted sum of its relative immunogenicity and the effective expansion rate as described in Equation (5.5). Since we have assumed that the effective expansion rate is the same across all CTL types, it follows that, using Equation (5.5), we can derive the relative immunogenicity for each individual epitope-specific CTL to be equal to its measured relative

[14]Recall that the effective expansion rate for CTL type k is given by $\phi_k = \nu_k / \mu_{4,k}$, where $k = 1, \ldots, p$. The assumption here means that this growth/death ratio ϕ_k is the same for all CTL types, although the individual proliferation and death rates may differ.

CTL abundance; that is, for each $k = 1, \ldots, p$,

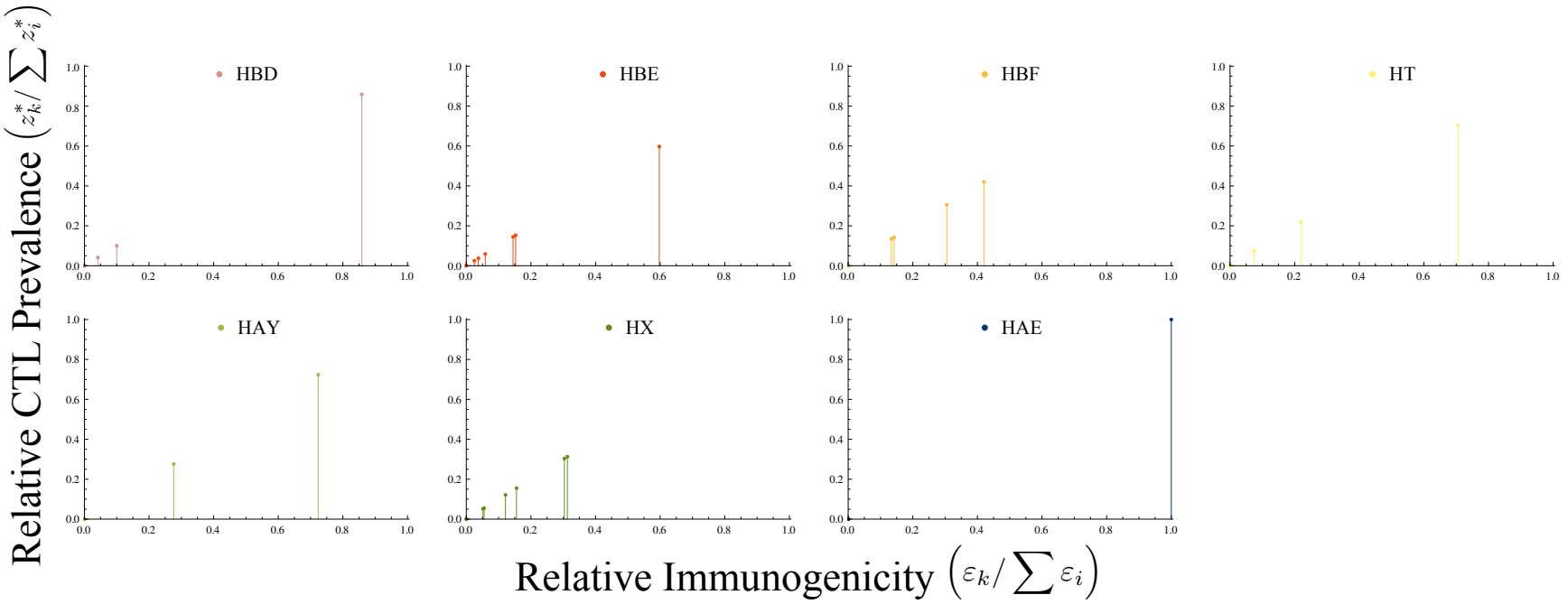
$$\underbrace{\frac{\varepsilon_k}{\sum_{j=1}^{p} \varepsilon_j}}_{\text{relative immunogenicity}} = \underbrace{\frac{z_k^*}{\sum_{j=1}^{p} z_j^*}}_{\text{measured relative CTL abundance}}. \tag{5.7}$$

Applying Equation (5.7) to the data listed in Table 5.6 with $p = 8$ different CTL types allows us to calculate the relative immunogenicities for the various epitopes, which taken together provides us with the immunogenicity profiles for each of the HTLV-I seropositive individuals considered by Goon et al. (2004a). We then plot the immunogenicity profiles for each patient in the form of list plot representations, shown in Figure 5.10, as we have done previously in Section 5.3.3 using simulated data. The corresponding values for the relative immunogenicities of each epitope are shown in Table 5.8. As we have discussed previously, this list plot representation helps us to visualise the relative distances in immunogenicities amongst the various epitopes. Thus, we are able to establish the immunogenicity profiles for all 10 HAM/TSP patients (Figure 5.10(a)) and 7 healthy asymptomatic carriers (Figure 5.10(b)). A combined plot of the immunogenicity profiles for all individuals from each of the two cohorts is shown in Appendix Figure C.5 for reference.



(a) Immunogenicity profiles for HAM/TSP patients in the data.

Figure 5.10: **(Caption on following page).**



(b) Immunogenicity profiles for the ACs in the data.

Figure 5.10: List plot representations of the immunogenicity profiles of HTLV-I-infected individuals as described by experimentally measured data from Goon et al. (2004a) for (a) HAM/TSP patients, and (b) ACs. The individual plots show the distribution of relative immunogenicities towards each of the eight epitopes (horizontal axis) and the relative abundances of their cognate CTL types (vertical axis). The immunogenicity profile for each HTLV-I seropositive individual is obtained by assuming that the effective expansion rate is the same for all CTL types and, using Equation (5.5) derived from model (5.1), determining that the relative immunogenicity for each epitope-specific CTL is equal to its measured relative abundance. The values that describe the immunogenicity profile for each HTLV-I-infected subject is shown for reference in Table 5.8. A combined plot of the immunogenicity profiles for all individuals from each of the two cohorts is shown in Appendix Figure C.5.

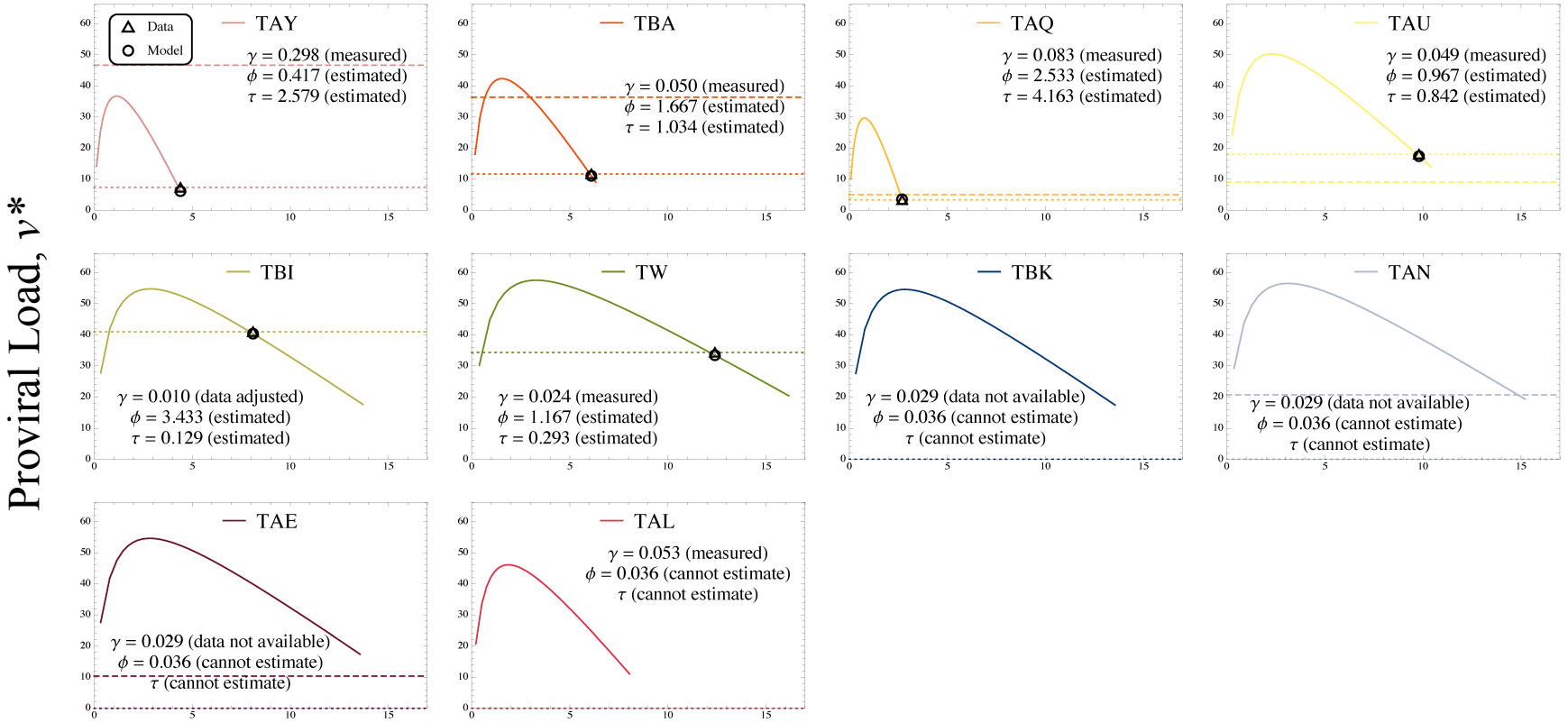
Clinical Status	Subject/ Patient	Relative Immunogenicity Towards Epitope							
		Tax13	Env	Rex	Tof	Rof	Gag	Pol	Pro
HAM/TSP	TAY	0.32	0.211	0	0.0581	0	0.113	0.297	0
	TBA	0.121	0.383	0	0.191	0.129	0	0.173	0
	TAQ	1.000	0	0	0	0	0	0	0
	TAU	0.135	0.0757	0.0384	0.0613	0.0669	0.0295	0.593	0
	TBI	0.734	0	0	0	0	0	0.266	0
	TW	0.829	0.123	0	0	0	0.0491	0	0
	TBK	0.588	0.247	0	0	0	0.0499	0.0645	0.0499
	TAN	0.780	0.0887	0	0	0	0	0	0
	TAE	0.856	0.0368	0	0.0233	0.0263	0.0249	0.0323	0
	TAL	0.842	0.116	0	0	0	0.0425	0	0
AC	HBD	0.859	0	0	0	0	0.0416	0.100	0
	HBE	0.597	0.145	0	0	0.0375	0.153	0.0589	0.0249
	HBF	0.420	0	0	0	0.142	0.134	0.305	0
	HT	0.706	0.0746	0	0	0	0	0.219	0
	HAY	0.724	0	0	0	0	0	0	0.276
	HX	0.313	0.0517	0	0.155	0.0551	0.121	0.304	0
	HAE	0	0	0	0	0	0	1.000	0

Table 5.8: Estimated immunogenicity profiles for each of the 17 HTLV-I seropositives considered by Goon et al. (2004a) according to the experimental data listed in Table 5.6 and using Equation 5.7. The distribution of the relative immunogenicities using our 'list plot' representation (as described in Section 5.2.4) corresponding to this table is shown in Figure 5.10.

We now turn to the data measurements of viral burden and CTL efficiency[15] for the subjects in the cohort under consideration (refer to Table 5.7, shaded rows), which we will use to estimate the remaining parameters. To examine viral detriment, we consider a similar graphical representation of viral burden as our usual one, where the activated infected proportion of all target cells (instead of all proviral cells) is along the horizontal axis and the proviral load is along the vertical axis (for example, see Figure 3.11 or Figures 5.4(c), 5.6(c), 5.8(c)). As mentioned previously, the experimental measurements for viral burden from Table 5.7 are directly analogous to, but not precisely the same as, the respective quantities that arise from our mathematical model, and the only difference would be in the scaling of the quantities, athough the exact scaling is not known. However, since the purpose of the current exercise is to demonstrate a proof of concept and not necessarily present a strict method, we will consider the values for the proviral load in the experimental data adjusted by our selected scaling factor of 0.3 $CD4^+$/PBMC as discussed and show how one could use our model to derive estimates for certain parameters. As we will see, even with this ad-hoc choice of scaling for the proviral load, the parameters that we can estimate still end up lying within biologically reasonable ranges, the implications of which we will discuss in more detail later.

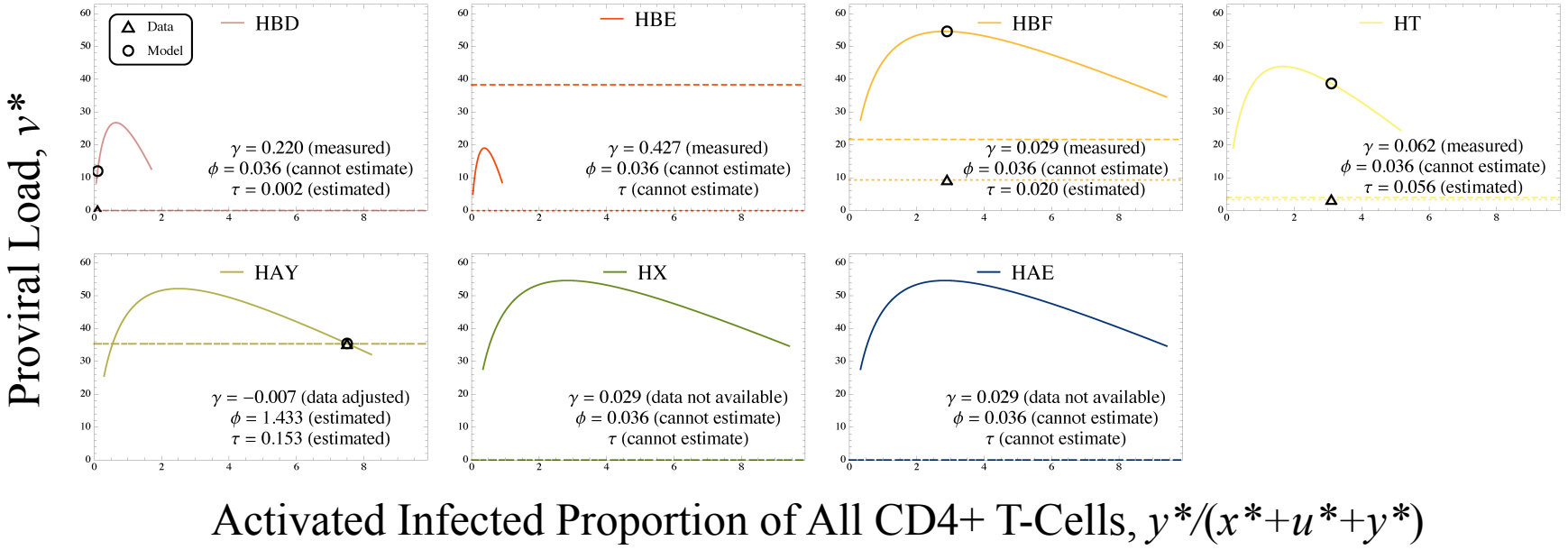
Figures 5.11(a)–(b) show the outcome of our parameter estimation using model (5.1). We also refer the reader to Table 5.9 for the measured and estimated values corresponding to both experimental data and our model results. We describe the procedure in more detail below. The plots combining the outcome of our parameter estimation for all HTLV-I-infected subjects in each of the two cohorts, HAM/TSP patients and ACs, are shown for reference in Appendix Figure C.6.

[15] We note here again that the measurements of the rates of cell-mediated lysis represent the total overall efficiency for the pool of CTLs — we do not have enough information to determine the individual lysis rates for each of the epitope-specific CTL types.



(a) Viral burden and parameter estimation in the subset of HAM/TSP patients.

Figure 5.11: Here we list the terminology used. Measured: Parameter value for the efficiency of HTLV-I-specific CTLs (γ) has been taken from experimental data. Data adjusted: The measured value for γ has been adjusted slightly because the original value was too small for the default algorithm to solve the system numerically (i.e. it was too close to 0). Data not available: The particular measurement for γ was not available, so we have substituted the parameter with the appropriate value from Table 3.1. Estimated: Estimated parameter value for either ϕ or τ by simple inspection and manual adjustment. Cannot estimate: The parameter estimation technique could not be done due to insufficient data. **(Caption continued on following page).**



(b) Viral burden and parameter estimation in the subset of ACs.

Figure 5.11: **(Continued caption from preceding page).** Comparison of viral burden or detriment as predicted by our multiple epitope model (5.1). Experimentally measured data are taken from the cohort of 17 HTLV-I-infected individuals considered by Goon et al. (2004a) consisting of (a) 10 HAM/TSP patients, and (b) 7 ACs. Matching the mathematical model to experimental data allows us to derive estimates for two key parameters: (i)the overall effecive expansion rate ϕ of CTLs, and (ii) the rate τ of spontaneous viral expression. To briefly outline the procedure, we first estimate, by inspection and manual fitting, the overall effective CTL expansion rate ϕ so that the curve describing viral detriment, namely how the active proportion of proviral cells varies with proviral load, generated using our model passes through the experimental data point. In the situations where ϕ cannot be estimated, we instead take median values from Table 3.1 as previously discussed (note: that here $\beta = 0.006$ mm^3/cell/day). Next, we can estimate the rate τ of spontaneous viral expression so that the magnitudes of the proviral load from our model matches those from the experimental data. The measured and estimated values for the various parameters and quantities related to viral detriment are shown in Table 5.9. △ Experimental data points. ◯ Estimated point using model (5.1). Solid line: Curve describing viral detriment as a parametric function of τ, the rate of spontaneous proviral expression, as predicted by the mathematical model. Dashed line: The level of the proviral load as reported in Goon et al. (2004a) (plotted for reference purposes only — we did not use this value for parameter estimation; see Table 5.6). Dotted line: The level of the proviral load as reported by Asquith et al. (2005a,b), adjusted by a chosen scaling factor of 0.3 CD4$^+$/PBMC to convert the measurement of the proviral load in terms of infected proportions of CD4$^+$ helper T-cells instead of PBMCs (wherever possible, we matched our model results to these experimental values; see Table 5.9). Combined plots of the above parameter estimation within each cohort, HAM/TSP patients and ACs, are shown in Appendix Figure C.6.

Clinical Status	Subject/ Patient	Measured Data from Table 5.7				Estimated Values From Model (5.1)			
		Rate of CTL Lysis Tax-Specific (γ) (per $CD8^+$ cell/day)	% Tax Expression ($Tax^+CD4^+/CD4^+$)	Proviral Load (% PBMC)	Adjusted Proviral Load (% $CD4^+$)	Rate of Effective CTL Expansion (ϕ)	Rate of Spontaneous Viral Expression (τ)	Tax^+ Proportion of $CD4^+$ T-Cells (%) $y^*/(x^*+u^*+y^*)$	Proviral Load (%) v^*
HAM/TSP	TAY	0.298	4.4	2.2	7.3	0.417	2.579	4.40	6.12
	TBA	0.050	6.1	3.5	11.7	1.667	1.034	6.10	11.05
	TAQ	0.083	2.7	1.0	3.3	2.533	4.163	2.69	3.51
	TAU	0.049	9.8	5.4	18.0	0.967	0.842	9.80	17.37
	TBI	0.003	8.1	12.3	41.0	3.433	0.129	8.10	40.37
	TW	0.024	12.4	10.3	34.3	1.167	0.293	12.40	33.44
	TBK	ND	ND	ND	N/A	CE	CE	CE	CE
	TAN	ND	ND	ND	N/A	CE	CE	CE	CE
	TAE	ND	ND	ND	N/A	CE	CE	CE	CE
	TAL	0.053	ND	ND	N/A	CE	CE	CE	CE
	Median[§]	0.050	7.1	4.5	14.8	1.417	0.938	7.100	14.210
	Mean[§†]	0.080	7.3	5.8	19.3	1.697	1.507	7.249	18.644
	Std Dev[§‡]	0.099	3.6	4.6	15.2	1.111	1.565	3.577	15.072
AC	HBD	0.220	0.1	0.001	0.003	CE	0.002	0.10	12.03
	HBE	0.427	ND	ND	N/A	CE	CE	CE	CE
	HBF	0.029	2.9	2.8	9.3	CE	0.020	2.90	54.58
	HT	0.062	3.1	1.0	3.3	CE	0.056	3.10	38.73
	HAY	-0.007	7.5	10.6	35.3	1.433	0.153	7.50	35.42
	HX	ND	ND	ND	N/A	CE	CE	CE	CE
	HAE	ND	ND	ND	N/A	CE	CE	CE	CE
	Median[§]	0.062	3.0	1.9	6.3	1.433	0.038	3.001	37.072
	Mean[§†]	0.146	3.4	3.6	12.0	1.433	0.058	3.402	35.188
	Std Dev[§‡]	0.179	3.1	4.8	16.0	N/A	0.067	3.057	17.559
Baseline Values		0.029	N/A	N/A	N/A	1.2	0.003	N/A	N/A

Table 5.9: **(Table caption on following page).**

Table 5.9: **(Table caption).** Measured and estimated values corresponding to viral burden and CTL properties (refer to Figure 5.11 for the graphs). Using experimentally measured data, namely the rate of Tax-specific CTL lysis, the percentage of $CD4^+$ T-cells expressing Tax (which is analogous to the active proportion of proviral cells), and the proviral load in terms of the proportion of Tax copies per 100 PBMCs (which is analogous to the proviral load in terms of proportion of infected target cells, which we obtain by scaling by a factor of 0.3 $CD4^+$/PBMC), we use model (5.1) to estimate the values of two parameters for each infected individual, specifically (i) the overall effective rate of CTL expansion, denoted by ϕ, and (ii) the rate of spontaneous proviral expression, denoted by τ, which is directly related to the fitness of the virus in terms of the basic reproduction number for viral infection. For infected individuals in the cohort under investigation for which we can estimate the effective rate of CTL expansion, we can compare this estimated rate to the baseline rate of 1.2, which is obtained by calculating the ratio $\phi = \nu/\mu_4$ using the baseline parameters as in Table 3.1. In the 7 subjects (6/10 HAM/TSP and 1/7 AC) for which a comparison can be made, we notice that the median and mean estimated values for the effective CTL expansion rate ϕ are slightly higher than the baseline value; however, no conclusions can be made about this result as individually, 4/7 subjects displayed higher estimated rates [highlighted in cyan (higher estimated ϕ)], and 3/7 displayed lower estimated rates [highlighted in lime (lower estimated ϕ)]. In contrast, with respect to the rate τ of spontaneous viral expression, which determines viral fitness, we find that both the median and mean estimated values for τ using our mathematical model is higher in HAM/TSP patients [highlighted in red (higher rate)], than in ACs [highlighted in (lower rate)], an observation which lends support to current experimental hypotheses that associates a high rate of proviral expression with the onset of disease. Baseline values for the parameters of interest here, from Table 3.1, are shown in the last row for reference. ND: No Data. N/A: Not Applicable. CE: Cannot Estimate. [§]Only calculated for values that are either measured or can be estimated. [†]Arithmetic mean. [‡]Sample standard deviation.

First, for those patients in the cohort with available data for the two measures of viral burden in Asquith et al. (2005a,b), which comprise 6/10 of the HAM/TSP patients and 4/7 of the healthy carriers, we plot the particular data point describing viral detriment for each HTLV-I-infected subject as a coordinate indicated by an open triangle $\triangle$, where the first coordinate describes the percentage of Tax^+ $CD4^+$ T-cells out of all $CD4^+$ T-cells ()i.e. the activated infected proportion of all $CD4^+$ T-cells), and the second coordinate is the proviral load in terms of infected $CD4^+$ T-cells (i.e. the adjusted proviral load, which is obtained by considering the reported value for the proviral load with respect to infected PBMCs and scaling it by a factor of 0.3 $CD4^+$/PBMC).

Next, for each patient, we plot the corresponding curve for viral detriment according to model (5.1) and select the overall effective CTL expansion rate in order for the system to achieve the correct balance of immune effectors so that the theoretical curve passes through the experimental data point, thereby providing us with an estimate of the parameter ϕ, which we are able to obtain for all 6/6 of HAM/TSP patients and 1/4 of ACs for which data are available (i.e. 6/10 of the HAM/TSP patients and 1/7 of ACs from the original cohort). Comparison of the estimated overall effective expansion rates of CTLs with the baseline rate of 1.2 (obtained by calculating the ratio $\phi = \nu/\mu_4$ using the baseline parameters

as in Table 3.1) reveals that the median and mean estimated values for the effective CTL expansion rate ϕ are slightly higher than the baseline value; however, on an individual level, out of the 7 subjects (6/10 HAM/TSP and 1/7 AC) for which a comparison can be made, 4/7 subjects displayed higher estimated rates and 3/7 displayed lower estimated rates (refer to the cyan- and lime-coloured highlighted entries in Table 5.9). We conclude that these estimated values for the overall effective rate of CTL expansion are in agreement with baseline values.

Lastly, we recall that since the curve describing viral burden using our mathematical model is obtained as a parametric function of the rate of spontaneous viral expression (τ), we can actually determine the value of this parameter for which the point on this curve matches the experimental data point. In particular, for the cohort under investigation, we select the value of τ so that the activated infected proportion of all $CD4^+$ T-cells between the experimental data and the model are in agreement. If we do not require the magnitude of the proviral load for each individual to simultaneously agree perfectly, this allows us to obtain an estimate for the spontaneous rate τ of viral expression for all 6/6 of the HAM/TSP patients[16] (Figure 5.11(a)) in addition to all 4/4 of the ACs (Figure 5.11(b)) for which data are available (i.e. 6/10 of the HAM/TSP patients and 4/7 of the ACs from the original cohort). The point on the graph describing viral burden corresponding to the estimated parameters is a coordinate indicated by an open circle ◯ (while the data point from experimental measurements, as mentioned previously, is indicated by an open triangle △). Examining the estimated values in closer detail, we discover that both the median and mean values for the rate, τ, of spontaneous viral expression in the subset of HAM/TSP patients are substantially higher (approximately 25 times for both metrics) than those for the subset of ACs (refer to the red- and yellow-coloured highlighted entries in Table 5.9). This observation lends support to experimental evidence demonstrating a strong association between the rate of spontaneous expression of the provirus and clinical status (Asquith and Bangham, 2007, 2008).

Remark 5.6. In the above representation of the experimental data, we do not incorporate the epitope specificities when building the immunodominance hierarchy for each HTLV-I seropositive individual, but instead, we distinguish amongst different CTL structures by the

[16] We note that there is very good agreement between the model and the data for the 6/6 HAM/TSP patients with respect to both measures of viral burden. However, this was true for only 1/4 ACs. For the remaining 3/4 ACs, although the model results matched exactly with the data with respect to the activated infected proportion of all $CD4^+$ T-cells, in contrast, there was a large discrepancy (specifically, an over-estimation) in the resulting predicted magnitude of the proviral load. This discrepancy has likely arisen because we have not distinguished HTLV-I sero-positive individuals by clinical status, i.e. HAM/TSP patient versus AC, whose baseline parameter sets, which influence the outcomes of the mathematical model and parameter estimation procedure, are bound to differ.

CTL rank order with respect to immunogenicity, e.g. the immunodominant CTL response in one infected individual may be directed towards Tax, whereas in another individual, the immunodominant CTL type might be targeted at a different viral protein such as Env.

Remark 5.7. Here, we have taken published experimental data and used our multiple epitope model (5.1) to estimate the average effective rate of CTL expansion and the rate of spontaneous viral expression for each infected individual as well as the corresponding resultant measures of viral burden, which we then compared to measured and estimated values in the literature using different methods. However, because all of the parameters are not exactly known for each individual, we have made several assumptions about their values, such as assuming the expansion rates for all CTL types within a host are equal to the average rate. Indeed, even in incorporating the experimentally measured rate of CTL lysis into our multiple epitope model (5.1), we have made the assumption that all CTLs directed against the various HTLV-I epitopes lyse their targets at the same rate. Although this may lead to a crude estimate, at best, of any quantities dependent on these values, it nevertheless provides us with a starting point for incorporating data into our mathematical framework and observing agreement or disagreement between experimental and estimated rates as well as qualitative trends. The purpose of this exercise at the present stage, therefore, is a proof of concept that our model could be used to estimate parameter values for quantities that either have not yet been, or otherwise cannot easily be, measured experimentally. An important next step would be to determine the appropriate scaling for each measurement in the literature so that the data values correspond precisely to the quantities represented in the mathematical model, which may then lead to a more accurate parameter estimation. Here we have chosen an ad-hoc scaling factor of 0.3 $CD4^+$/PBMC; however, the exact scaling is now known, is bound to vary across different individuals as well as in the same individual over time, and is subject to a great deal of uncertainty. To evaluate the validity of our parameter estimates, it will therefore be necessary to determine the corresponding experimentally measured quantities and see how they compare with our theoretically derived results.

Remark 5.8. In this section, we have estimated the parameters for the overall effective CTL expansion rate (which is a measure of CTL fitness) and the rate of spontaneous viral expression (which is a measure of viral fitness as it determines the basic reproduction number for viral infection R_0) through brute force trial and error. Specifically, we used simple inspection and manual adjustment to approximate the parameter values that resulted in a 'close' match between the model and the given experimental data, where by 'close', we mean that the measured (i.e. from the data) and estimated (i.e. from the model) values

of the proviral load for the patients generally agreed to two significant figures). Although our parameter estimation method has served its intended purpose of demonstrating a proof of concept that our model can be integrated with data to produce estimates on quantities that may be difficult to determine experimentally, we recognise that this rough approach is preliminary and requires more work to improve its accuracy and reliability. For instance, we have not yet performed a thorough sensitivity analysis of the mathematical model, nor have we established the uniqueness of our parameter estimation method. There is also a great deal of uncertainty in the data which we would like to address in the future as well. Whereas in the case of the 6/6 HAM/TSP patients for which viral burden data was available, we were able to determine both parameters of interest, namely ϕ and τ, which yielded nearly perfect matches between our multiple epitope model and the data, in contrast, we were able to determine both parameters for only 1/4 ACs. At this stage, it is unclear whether, firstly, there always exists a combination of both parameters that yield the given data (e.g. for the ACs, we could not find such a pair, but that does not mean it does not exist) and, secondly, if the parameters that we can estimate are unique (e.g. for the HAM/TSP patients, there could be an alternative combination of parameters that yield the same data). Nevertheless, we are aware that there exist systematic methods for parameter estimation and matching experimental data to mathematical models that are more rigorously defined. This leads to several active areas of research such as identifiability, uncertainty analysis, and Bayesian inference, which employ powerful mathematical and statistical tools that are beyond the current scope of this thesis.

5.5 Discussion of Main Conclusions Drawn from the General Multiple Epitope Model

The focus of this chapter has been to develop a consistent framework to understand the concept of immunodominance, which helps us to unravel the mechanisms by which the cellular immune response acts against a pathogen expressing multiple viral antigens simultaneously. Although our investigation is in the context of persistent infection by HTLV-I, the concepts we have developed can be applied more generally to any host-pathogen system where a homogeneous pathogen population harbouring a number of distinct loci can elicit an array of epitope-specific immune responses.

In particular, we generalise the basic two-epitope Tax/HBZ model (4.1) to incorporate the simultaneous expression of any arbitrary number of epitopes, yielding the multiple epitope model (5.1). A key feature of the generalised multiple epitope model (5.1) is the consideration of a wide range of HTLV-I-specific CTL responses targeting the array of viral proteins that are expressed by actively infected cells. By considering the properties for

each CTL, we can construct an immune profile for each individual and explore the way in which different immune profiles influence the measures of viral burden, namely proviral load and the active proportion of infected cells. Our investigations have allowed us to draw the following conclusions based on our mathematical modelling approach.

Q1. What can our model say about immunodominance in the context of chronic HTLV-I infection and the role of CTL structure?

▷ **We have demonstrated using our multiple epitope model** (5.1) **that immunodominance, and the organisation of an immunodominance hierarchy, is the outcome of two factors: (i) epitope immunogenicity, and (ii) the effective expansion rate of the corresponding CTL type. Within a single individual, although effective CTL expansion rates are believed to remain stable over time, different immunogenicity profiles (defined by the set of relative immunogenicities of epitopes at a given time) can give rise to a broad range of CTL population structures, which differentially affects viral burden. We find that effective control of the virus infection is the consequence of an immunogenicity profile that encourages the collective dominance of efficient CTL types that expand rapidly and are highly immunogenic. Lastly, our results show that the immunodominant response is not necessarily the most effective one at reducing viral detriment.**

In terms of CTL abundance, we have shown that the prevalence of a particular CTL type, which is equivalent to its degree of dominance, is determined by a combination of only two factors: the relative immunogenicity of the CTL epitope, and the effective rate of CTL expansion. The immunogenicity profile refers to the distribution of relative immunogenicities of each of the various epitopes. Epitopes can be listed in a rank order according to ascending immunogenicity and this rank order is believed not to change over the course of chronic infection. For instance, during chronic HTLV-I infection among both asymptomatic carriers and HAM/TSP patients, epitopes from the viral Tax protein (which is expressed early on in the virus replication cycle) are highly immunogenic and are also immunodominant, eliciting the largest CTL response, and continue to do so throughout the course of infection (Bangham et al., 1999; Cook et al., 2013; Goon et al., 2004a; Kattan et al., 2009). Other subdominant immune responses, such as those directed towards epitopes from the proteins Pol, Env, Gag, or the anti-sense gene product HBZ, may be present, but remain poorly immunogenic

and unable to surpass Tax in terms of immunogenicity[17] (Goon et al., 2004a).

Within an individual, it has been observed in clinical data that a change in the prevalences of the various effector subpopulations can either preserve or shift the structure or immunodominance hierarchy of the CTL response, i.e. the ranking of CTL types in order of their relative abundances. Indeed, to take a concrete example, in the experimental study by Daenke et al. (1996), it was observed that the immunodominant epitope (specifically, the X19 peptide from the C-terminal end of the Tax protein) remained immunodominant throughout a five month period in one infected individual (TZ). By contrast, in a different HTLV-I seropositive individual (HM), the immunodominant epitope was switched from XN11 to XN3 (these are two different peptides from the N-terminal end of the Tax protein) over a period of time (Daenke et al., 1996).

In our modelling investigations, we have reasoned that within the same host, intrinsic CTL characteristics, including the rates of CTL proliferation and death that define CTL expansion, are unlikely to vary drastically over time, with the exception of a change in phenotype of the infection, for example, if the infected individual experiences a sharp decline in health, as with the development of pathological disease from a previously asymptomatic state. Thus, if all immune responses retain their respective effective expansion rates, the only other possibility for a change in CTL structure to occur in our model is if there is variation in the relative immunogenicities among the epitopes. As variation in the immunogenicity profile of an individual is likely to occur over time due to the constantly evolving and changing nature of the within-host environment, such a scenario is immunologically feasible. After all, the interactions that take place within an infected individual are not static, but rather give rise to a dynamic and ever-changing battle between the virus and host immunity and, essentially, any factor that can influence the immunogenicity of even a single epitope or its CTL response can effect a substantial change in the immunogenicity profile of an individual. For instance, the emergence of a previously undetectable epitope-specific CTL type, or the loss of recognition of an epitope, can affect the relative

[17] Exceptionally, there is also evidence that a change in the rank order of epitopes with respect to immunogenicity may be accompanied by a drastic transformation of the infection phenotype (e.g. the development of a pathological condition). For example, while the transcriptional transactivator protein Tax plays a central role in the establishment of chronic HTLV-I infection and is detectable in all HTLV-I-seropositive asymptomatic carriers, it has been observed in fresh leukaemic isolates derived from patients with the T-cell malignancy adult T-cell leukaemia (ATL) that Tax is not readily expressed in 60% of ATL cases (Matsuoka, 2005). By contrast, the reverse-strand viral gene product HBZ is poorly detected in asymptomatic carriers, but can be identified in 100% of ATL patients (Satou et al., 2006; Saito et al., 2009). This observation suggests a significant shift not only in the immunogenicity profile of an HTLV-I seropositive individual between non-malignant and malignant infected states, but also in the immunogenicity rank order of epitopes.

immunogenicities, and hence the immunogenicity profiles, of the anti-HTLV-I CTL response within the same host across two different time points. Parker et al. (1994) observed both of these scenarios experimentally. Additionally, cross-immunity arising from non-HTLV-I infections can sometimes lead to recognition by, and subsequent expansion of, anti-HTLV-I CTLs, which would affect the immunodominance hierarchy by changing the structure of the CTL population. Because the immunogenicities of epitopes are likely to change throughout the course of the infection, we investigated this idea in slightly more detail using our multiple epitope model framework.

We demonstrated that considering only differences in the immunogenicity profiles can reproduce both of the aforementioned behaviours, that is, either preserving or changing the structure of the HTLV-I-specific CTL response, even when the rank order of the CTL epitopes with respect to immunogenicity is maintained. This is due to the fact that despite taking into account the maintenance of such CTL rank order, the relative immunogenicities of each epitope can still vary a great deal, giving rise to a variety of different immunogenicity profiles. Indeed, pairwise comparison of immunogenicity profiles can result in either the maintenance of CTL immunodominance hierarchy with possibly varying relative prevalences (for example, compare IP-2 with IP-3 in Figure 5.3), or a complete shift in the immunodominant CTL response (for example, compare IP-1 with IP-2 in Figure 5.3).

Furthermore, we showed that the composition of the CTL population brought about by distinct immunogenicity profiles can differentially influence the viral burden due to the infection. We discovered that immunogenicity profiles which favoured a higher proportion of the immune cell population being composed of efficient CTL types led to more effective control of viral detriment. This could be achieved by rapid expansion or by high immunogenicity of efficient CTL types, whether or not the immune response is more heterogeneous or more homogeneous. Along these lines, our model results demonstrated that the effective control of viral detriment is not necessarily achieved by the action of the immunodominant CTL response. Conversely, poor control of the virus population was the result of immunogenicity profiles that favoured immunodominance of an inefficient CTL response or a slow expansion of efficient CTL types.

Q2. How can our mathematical framework be used to relate to experimental data?

▷ **Using published experimental data in the context of our multiple epitope modelling framework, we are able to estimate values for the immunogenicity profile as well as two key parameters, which have not been experimentally measured, in the majority of HTLV-I seropositives in the cohort under investigation. We discover that our estimated parameter values for the overall effective rate of CTL expansion are generally in agreement with baseline values. More significantly, the outcome of our parameter estimation for the rates of spontaneous proviral expression, which is a measure of viral fitness, reveal that the virus population in HAM/TSP patients exhibits a greater rate of viral expression, on average, than healthy asymptomatic carriers, thus providing support for a recent experimental hypothesis that a high degree of viral expression is associated with the progression of HTLV-I-linked disease.**

We have collated and incorporated experimental data for a total of 17 HTLV-I seropositive individuals (namely, 10 HAM/TSP patients and 7 ACs) into our mathematical framework from three primary sources: Goon et al. (2004a) and Asquith et al. (2005a,b). Using model (5.1), we are able to determine the immunogenicity profiles for all infected subjects, and to derive estimates for the rate of spontaneous proviral expression, as well as the overall effective rate of CTL expansion, for the majority of patients for which data were available.

We emphasise two main conclusions that arise from the results of our parameter estimation. First, we discovered that among all of the HTLV-I-infected individuals in the cohort under investigation for which we were able to successfully estimate the appropriate parameters, the overall effective CTL expansion rates were in general agreement with the baseline value, thus supporting the validity of our parameter estimation technique.

Second, and more significantly, we observed that the average estimated rate of spontaneous proviral expression was significantly higher in the subset of HAM/TSP patients than in the subset of healthy ACs, suggesting that a higher rate of spontaneous viral expression is linked to the presence of HTLV-I-associated disease. This hypothesis has recently been tested using experimental methods which demonstrated a strong association between a high rate of spontaneous viral expression, which is believed to be an

intrinsic characteristic of the virus infection within a given host, and pathological clinical status, namely the neurological and inflammatory disease HAM/TSP (Asquith and Bangham, 2007, 2008).

Q3. How does our modelling approach relate to similar approaches?

▷ **We have developed the first within-host multiple epitope mathematical model of HTLV-I infection. Other authors have considered the simultaneous expression of multiple antigens in different viral infections, notably HIV. One crucial difference is that our framework allows for the co-existence of different epitope-specific CTL responses and does not depend on viral mutation leading to immune escape in order to establish an immunodominance hierarchy. We exploit these features to characterise effective host immunity.**

The phenomenon of immunodominance has been studied by both experimentalists and theoreticians alike (Akram and Inman, 2012; Althaus and Boer, 2008; Nowak and Bangham, 1996). Nowak and Bangham (1996) examined immunodominance using a predator-prey model to describe the interactions between epitope-specific CTLs and the virus population, primarily in the context of HIV infection. The authors concluded that persistent infection by a homogeneous virus population should be expected to yield immunodominance of a single epitope. Moreover, by introducing heterogeneity in the virus population via the consideration of antigenic variation, there can arise antigenic oscillations in which the abundance of different virus variants, as well as the dominance of immune effectors, fluctuate throughout time as a consequence of changing immune selection acting on the diverse virus pool. Our model conclusions agree with those of Nowak and Bangham (1996) in establishing that immunodominance of a CTL response is a consequence of epitope immunogenicity and viral replication rate. However, the model by Nowak and Bangham (1996) precludes the co-existence of multiple epitope-specific CTL responses unless mutational events are taken into consideration. By contrast, our multiple epitope modelling approach allows for such a scenario even when no mutation is present and there exist functional differences amongst the various epitope-specific effectors present in the CTL repertoire.

Althaus and Boer (2008) developed a computational model to characterise within-host infection by HIV/SIV focussing on the dynamics of immune escape arising from escape and compensatory mutations. The authors concluded that a greater degree of immunodominance leads to poorer control of viral replication. The results from our model (5.1) are consistent with those of Althaus and Boer (2008). However, one

difference is that our approach incorporates different cell-mediated lysis rates for each of the various CTL types, and what we find is that although a more heterogeneous CTL pool is generally more effective at controlling viral abundance, our model also demonstrates that effective immune surveillance is possible even when a single CTL type strongly dominates host immunity, provided that this particular immune response is adequately efficient in eliminating infected target cells.

Q4. What are the advantages of our mathematical modelling framework?

▷ **We have been careful to develop a biologically meaningful and mathematically rigorous within-host model of HTLV-I that captures the multiple epitope nature of the host-virus interaction and can be interpreted in the context of experimental observations.**

By characterising the global dynamical properties, we can be certain that the model is well-posed and that we understand all possible behaviours of the system. Moreover, we have demonstrated that taking into consideration the multiple epitope nature of host-virus interactions allows us to incorporate more detailed data sets which can help estimate key parameters that have not yet been measured or, perhaps in some instances, cannot be measured. We estimated these parameters in a cohort of HTLV-I seropositive individuals and drew conclusions from the results which match with current experimental hypotheses. Although our model has been formulated in the context of chronic infection by HTLV-I, the focus in this chapter has been on the immunological phenomenon of immunodominance and the concept of an immunodominance hierarchy. More generally, our mathematical framework can be applied to a wider range of host-pathogen invasions that exhibit immunodominance.

5.6 Final Remarks on the General Multiple Epitope Model

Immunodominance refers to the frequently observed phenomenon during chronic infection in which a single epitope (called the immunodominant epitope) elicits a disproportionately large CTL response compared to the other virus targets (known as the subdominant epitopes). Immunodominance is commonly, though not always, observed in a number of viral and bacterial infections, including the human immunodeficiency virus type 1 (HIV-1), influenza A, West Nile cirus, Listeria, human cytomegalovirus (CMV), hepatitis B virus (HBV), hepatitis C virus (HCV), and Epstein-Barr virus (EBV) (Akram and Inman, 2012; Frank, 2002).

In this chapter we have developed a novel framework to examine the host-virus interactions that take place when any arbitrary number of viral antigens are expressed simultaneously, which naturally implies the existence of a broad array of corresponding epitope-specific CTL responses. As a result, the anti-viral cellular immune response is made up of a diverse array of CTL types, each characterised by a distinct set of characteristics, including the effective rate of expansion, the immunogenicity of its cognate epitope target that stimulates CTL proliferation, the rate of cell-mediated elimination of infected cells, and the relative prevalence of that particular CTL type which creates a CTL immunodominance hierarchy or list of CTL responses from least abundant to most abundant. Due to these differences among the various CTL types, it is often the case that effective control of viral burden is driven primarily by the action of a small subset of the entire CTL repertoire. Notably, we determined that an immune response that is effective at controlling viral burden, with respect to the proviral load and activated proportion of proviral cells, is one that is comprised of CTL types that are efficient at killing infected cells, expand rapidly, and target epitopes that are sufficiently immunogenic to elicit a large enough response that is able to make an impact on the proviral population.

However, in our investigation, we have only considered heterogeneity in the CTL response and not in the virus population. In the following chapter, we will take a step towards an alternative extension of the two-epitope Tax/HBZ model (4.1) which incorporates the role of antigenic variability in the HTLV-I system.

Chapter 6

A Multi-Locus Approach With Antigenic Variation: The Multi-Locus Model

The two-epitope Tax/HBZ model (4.1) that we developed in Chapter 4 incorporated an important feature of HTLV-I-infected cells (and more generally of cells infected by a virus), namely that proviral activation is often accompanied by the simultaneous expression of multiple viral proteins, Tax and HBZ in our case. As we have mentioned, we have identified two different extensions of the Tax/HBZ model (4.1), each of which is aimed at elucidating a certain aspect of the host-virus interaction. In Chapter 5, we have generalised the two epitope model to include any arbitrary number of epitopes. Indeed, since the HTLV-I genome encodes a handful of viral proteins and not just Tax and HBZ, the multiple epitope model (5.1) allowed us to better understand how the existence of multiple viral gene products expressed simultaneously by a single infected target cell, the relative immunogenicities of such proteins, and the differences in the response of the adaptive immune system, can explain the phenomenon of immunodominance as well as define various types of CTL structures in the pool of epitope-specific anti-HTLV-I effector cells which can differentially impact the viral population.

Next, we consider an alternative extension of the two-epitope Tax/HBZ model (4.1) that focusses on a different, but equally important, biologically realistic aspect of the virus that has not yet been considered in mathematical models of HTLV-I infection. In particular, this alternative model extension is motivated by experimental evidence for the existence of naturally occurring variants in both Tax and HBZ (Cavanagh et al., 2006; Gaudray et al., 2002; Kubota et al., 2007; Niewiesk et al., 1995), a feature known as antigenic diversity or variability. Coupled with the inherent diversity presented by host immunity, such antigenic variability implies a complex interplay between the virus and the host. Our goal, therefore, is to examine how heterogeneity in the two interacting populations can shape the overall

dynamics of the infection. In this chapter, we bring together the conceptual framework we have been building in the previous chapters to develop a generalised model incorporating antigenic variability at multiple epitopes.

There are important biologically relevant aspects that can be explored by taking into account the multi-locus nature of HTLV-I, namely the heterogeneity in the population structure of virus-infected target cells. Moreover, the high specificity of the host cellular immune response to a particular variant of a viral antigen means that there is likely to be a large degree of diversity and structure present in the CTL population. We apply the results of our model to try and gain insights on the long-term persistence of HTLV-I infection, as well as to the pathogenesis of HTLV-I-associated disease.

6.1 Motivation

It is widely believed that a high proviral load is the principal factor contributing to disease manifestation. However, while asymptomatic carriers tend to carry a lower equilibrium proviral load, they share a very broad range of overlap with symptomatic carriers (Asquith and Bangham, 2007; Mosley et al., 2005; Nagai et al., 1998). Even though there is a large degree of diversity in viral strains, no single HTLV-I strain is known to be directly associated with disease outcome (Daenke et al., 1990; Kinoshita et al., 1991; Mosley et al., 2005; Niewiesk and Bangham, 1996). Nevertheless, studies have demonstrated that healthy HTLV-I seropositive individuals harbour virus populations displaying wider variation at the protein level than do patients with HAM/TSP (Niewiesk et al., 1994). A study on HTLV-I seroprevalence in two geographically different, but ethnically similar, communities has also provided evidence that a high degree of HTLV-I sequence conservation is associated with ATL (Moynet et al., 2001). Furthermore, the recent characterisation of infected T-cell clones distinguished by their unique sites of proviral integration has shown that the proviral population in infected patients with malignant disease (the various manifestations of ATL) is comprised principally of a relatively small handful of clone types that exist in large abundances and account for major proportions of the virus population (Gillet et al., 2011): for example, in one patient with chronic leukaemia, more than 80% of the HTLV-I-infected target cell population was made up of a single, abundant clone.

The main motivation behind developing a multi-locus model of HTLV-I infection is to examine how variability in the virus and the host influences the dynamics between HTLV-I and host immunity. Specifically, we focus on the way in which these interactions contribute to the structure and abundance of both the virus and anti-HTLV-I CTL populations. Using the more biologically realistic multi-locus approach raises several issues related to how

antigenic variability can aid viral persistence and contribute to disease manifestation. We consider the following questions that we would like to investigate using this extended model:

(1) How are the dynamics and outcomes of persistent infection by HTLV-I influenced by the extent of homogeneity or heterogenity in the proviral population? In particular, what effects do the relative sizes of each virus variant or clone have on the course of infection?

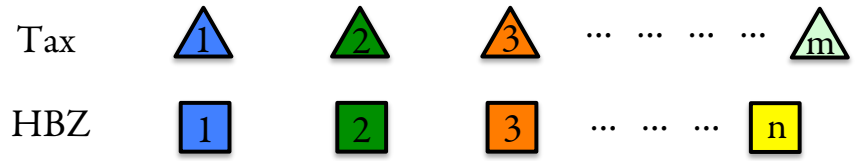
(2) Which is worse for the host: An infection that is largely homogeneous, or one that is antigenically diverse? If we know the particular structure of the proviral load for an infected individual who is undergoing treatment, must we employ different strategies depending on the degree of heterogeneity of the virus clone structure?

(3) What sort of similar questions can we raise with respect to the degree of heterogeneity and relative abundances present in the population of antigen-specific anti-HTLV-I CTLs.

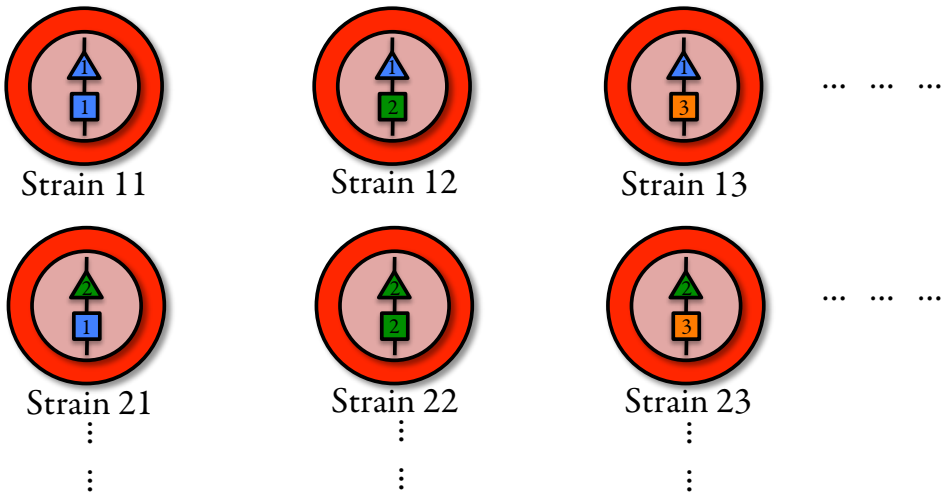
6.2 The General Multi-Epitope Model Incorporating Antigenic Variability at Multiple Epitopes Simultaneously: m Variants in Tax, n Variants in HBZ

The inherent heterogeneity of the HTLV-I genome acting in parallel to the diversity of associated host immune responses naturally suggests the use of a multi-locus modelling framework, which has been extensively developed in the context of immune selection on antigenically diverse infectious pathogens (Gupta et al., 1996, 1998; Gupta and Anderson, 1999; Gupta and Maiden, 2001), including the malaria parasite *Plasmodium falciparum* (Gupta et al., 1994; Gupta, 2005; Recker et al., 2004), the bacterium *Neisseria meningitidus* (Buckee et al., 2008; Callaghan et al., 2008), influenza (Recker et al., 2007; Wikramaratna et al., 2013), and the human immunodeficiency virus type 1 (HIV-1) (Nowak et al., 1995b,a).

Generally, there can exist any number of antigenic variants at each viral epitope and, from a biological point of view, this depends on factors such as whether variability in the epitope is subject to functional or conformational constraints that promote epitope conservation or diversity. Nevertheless, any given single proviral cell is capable of expressing just one variant of Tax and one variant of HBZ. Moreover, the combination of a certain Tax variant and a certain HBZ variant defines a distinct virus variant or clone as the antigenic variability in Tax and HBZ are presumed to be independent of one another. A diagrammatic illustration is shown in Figure 6.1.



(a) There exist naturally occurring variants in each of the two proteins.



(b) A virus strain is uniquely defined by a combination of each of its epitopes.

Figure 6.1: A multi-locus approach to modelling HTLV-I infection involves the consideration of antigenic variability at multiple epitopes simultaneously. (a) We suppose that there exist m distinct variants of the Tax protein and n distinct variants of the HBZ protein. (b) A virus strain ij is defined by the unique combination of Tax variant i and HBZ variant j, where $1 \leq i \leq m$ and $1 \leq j \leq n$.

Meanwhile, the anti-HTLV-I host immune response is highly specific and is maintained by the presence of expressed viral antigens. Despite the simultaneous expression of multiple viral protein variants (one at the Tax epitope and one at the HBZ epitope) in an activated proviral cell, only one epitope-specific CTL response may be stimulated at a time. The extent to which one epitope is recognised over another by corresponding host CTLs is therefore a measure of the immunogenicity of that particular epitope. A schematic of the key interactions is illustrated in Figure 6.2.

6.2.1 Formulation of the General Multi-Locus Model

We describe the above ideas in mathematical terms. Let m and n be positive integers. Suppose that there are m different variants in the genomic sequence of Tax and n different variants in that of HBZ. In the following, we define a virus 'strain' ij to refer to an infected target cell possessing a specific variant for each of the two major HTLV-I proteins, Tax variant i (Tax_i) and HBZ variant j (HBZ_j), where $1 \leq i \leq m$ and $1 \leq j \leq n$.

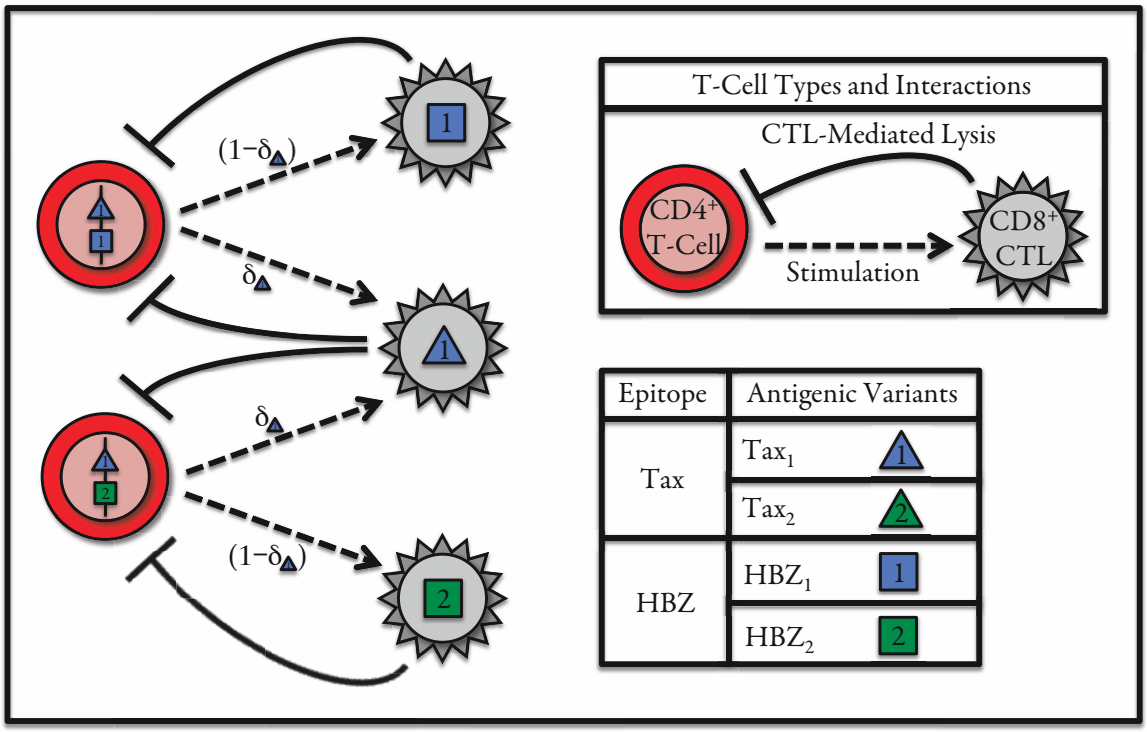


Figure 6.2: A schematic representation illustrating the key interactions of the multi-locus model (6.1). Each strain simultaneously expresses multiple viral antigens, one at each epitope (Tax and HBZ), and can therefore be recognised by multiple $CD8^+$ CTLs that are specific to one antigen or the other. However, any single actively infected target cell ($CD4^+$ helper T-cell) may stimulate only one CTL response at a time. Meanwhile, a single variant-specific CTL type is able to suppress multiple different strains as long as each of these strains share the same variant of the particular epitope. The degree of immunogenicity towards antigenic variant Tax_i at the Tax epitope, denoted by δ_i, determines the extent to which the Tax-specific or HBZ-specific CTL responses interact with the given strain.

The pool of healthy $CD4^+$ helper T-cells $\big(x(t)\big)$ is susceptible to infection by all virus strains. Transmission by strain ij, which encodes the viral protein variants Tax_i and HBZ_j, results in infected target cells that can exist in two possible states depending on the transcriptional activity of the provirus: latently infected cells $\big(u_{ij}(t)\big)$ and actively infected cells $\big(y_{ij}(t)\big)$. The host immune response is partitioned into two compartments: CTLs that are specific to the Tax_i variant at the Tax epitope $\big(z_i(t)\big)$, and CTLs that are specific to the HBZ_j variant at the HBZ epitope $\big(s_j(t)\big)$. We therefore consider the T-cell populations for

$i = 1, \ldots, m$ and $j = 1, \ldots, n$:

$$
\begin{aligned}
x(t) &: \text{healthy (uninfected) CD4}^+ \text{ helper T-cells,} \\
u_{ij}(t) &: \text{latently infected CD4}^+ \text{ helper T-cell variant } ij \text{ (with Tax}_i \text{ and HBZ}_j\text{),} \\
y_{ij}(t) &: \text{actively infected CD4}^+ \text{ helper T-cell variant } ij \text{ (expressing Tax}_i \text{ and HBZ}_j\text{),} \\
z_i(t) &: \text{CD8}^+ \text{ CTLs specific to Tax}_i\text{,} \\
s_j(t) &: \text{CD8}^+ \text{ CTLs specific to HBZ}_j\text{.}
\end{aligned}
$$

The general multi-locus mathematical model for the within-host dynamics of HTLV-I infection that incorporates antigenic variability in both epitopes is represented by the following $(2mn + m + n + 1)$-dimensional non-linear system of ordinary differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \lambda - \beta x \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij} - \mu_1 x, \\
\frac{\mathrm{d}u_{ij}}{\mathrm{d}t} &= (\beta x + r) y_{ij} - (\tau_{ij} + \mu_2) u_{ij}, \quad \text{for } i = 1, \ldots, m \text{ and } j = 1, \ldots, n, \\
\frac{\mathrm{d}y_{ij}}{\mathrm{d}t} &= \tau_{ij} u_{ij} - \gamma y_{ij} z_i - \gamma' y_{ij} s_j - \mu_3 y_{ij}, \quad \text{for } i = 1, \ldots, m \text{ and } j = 1, \ldots, n, \\
\frac{\mathrm{d}z_i}{\mathrm{d}t} &= \nu \delta_i \sum_{j=1}^{n} y_{ij} - \mu_4 z_i, \quad \text{for } i = 1, \ldots, m, \\
\frac{\mathrm{d}s_j}{\mathrm{d}t} &= \nu' \sum_{i=1}^{m} (1 - \delta_i) y_{ij} - \mu_5 s_j, \quad \text{for } j = 1, \ldots, n.
\end{aligned}
\tag{6.1}
$$

Because the above model considers m variants in Tax and n variants in HBZ, we sometimes refer to model (6.1) as the '$m \times n$' model. Many of the underlying processes of the infection dynamics are the same as those represented by the basic two-epitope Tax/HBZ model (4.1) and hence do not require detailed explanation. However, what is different here is the refinement of the anti-HTLV-I CTL response to reflect specificities to a particular variant of the two HTLV-I proteins at a time, Tax or HBZ but not both.

Healthy new CD4$^+$ helper T-cells are produced at a constant rate λ. All strains are assumed to have the same coefficient of infectious transmissibility β and rate of mitotic division r. Newly infected target cells silence viral expression and become latently infected. For each strain ij, there is a characteristic rate τ_{ij} at which infected target cells that are transcriptionally latent spontaneously become re-activated. The simultaneous expression of both viral antigens that accompanies spontaneous proviral activation means that these actively infected cells stimulate, and are targeted by, two types of anti-HTLV-I CD8$^+$ CTL responses. We assume that the two distinct CTL types, Tax-specific and HBZ-specific, proliferate at rates ν and ν', respectively, and may be functionally dissimilar in their effector

responses to each of the two epitope targets; namely, we denote the respective rates for Tax-specific and HBZ-specific CTL-mediated lysis by the parameters γ and γ'. A key feature of model (6.1) that we expand from the previous Tax/HBZ model (4.1) is that activated proviral cells cannot simultaneously initiate the expansion of both types of CTLs. We represent this aspect of the host-virus interaction by the parameters δ_i, where $0 \leq \delta_i \leq 1$ for $i = 1, \ldots, m$, which represents the immunogenicity of the Tax_i variant[1]. Then, the fraction $\delta_i y_{ij}(t)$ of fully activated proviral cells infected by strain ij stimulates the expansion of CTLs specific to Tax_i, whereas the remaining fraction $(1 - \delta_i)y_{ij}(t)$ stimulates the expansion of CTLs specific to HBZ_j. It is assumed that all actively infected proviral cells, which express one variant of Tax and one variant of HBZ, must stimulate exactly one of the two antigen-specific CTL responses. Large values of δ_i therefore correspond to a strong stimulation of Tax_i-specific CTLs and a weak stimulation of HBZ_j-specific CTLs. The reverse is true for small values of δ_i.

Lastly, healthy, latently infected, and actively infected target cells are removed from the system at linear rates μ_1, μ_2, and μ_3, respectively, whereas Tax-specific and HBZ-specific CTLs have respective removal rates μ_4 and μ_5. All parameters are assumed to be positive, with the exception of δ_i, which lies in the closed interval between 0 and 1.

Remark 6.1. Immunologically speaking, there is a distinction between a virus-infected clone and a virus strain. Whereas a virus strain represents a particar genetic variant of a virus genome, a clone refers to the genetically identical progeny of a virus-infected parent cell. In this way, many distinct clones can be and, in fact, are likely to be, infected with the same viral strain. Our definition of a virus 'strain' represents an infected target cell possessing a particular combination of one Tax variant and one HBZ variant (which are the product of different genes encoding each protein) and thus corresponds to the aforementioned virological notion of a strain.

6.2.2 Comments About Infection by Multiple Strains and Defining Measures of Strain and CTL Fitness

One of the implications of infection by multiple strains of a pathogen is a broadening of the definition of what constitutes a chronic infection. Indeed, a persistent infection can be caused by the presence of a single viral strain, or by a combination of multiple co-existing

[1]Recall that we have previously defined epitope immunogenicity in Chapter 4 to be the rate at which the particular epitope induces the proliferation of CTLs directed against it. In the special case, $m = n = 1$, where there is no antigenic variability in either the Tax or HBZ epitopes, this particular feature of simultaneous expression of multiple epitopes is captured by a single parameter δ, with $0 \leq \delta \leq 1$, and corresponds precisely to the concept of immunogenicity which we developed in the basic two-epitope Tax/HBZ model (4.1).

strains. We will therefore define chronic infection to mean the long term persistence of at least one viral strain.

As with the models without the inclusion of antigenic variability (e.g. the baseline model (3.1) and the Tax/HBZ model (4.1)), the fitness of a particular viral strain ij is directly related to the basic reproduction number for viral infection, also commonly known as the basic reproductive ratio, which we denote by $R_{0,ij}$. The quantity $R_{0,ij}$ is dependent upon three key factors: (i) the rate of infectious proviral transmission (β), (ii) the rate of selective mitotic division of proviral-expressing infected cells (r), and (iii) the rate at which latently infected target cells spontaneously express viral antigens (τ_{ij}). The form of $R_{0,ij}$ as in Equation (6.2) is analogous to the one derived for a single strain in previous models (refer to Appendix A.1 for various methods commonly used to calculate an expression for the basic reproduction number), and is given by:

$$R_{0,ij} = \frac{\tau_{ij}}{\mu_3(\tau_{ij}+\mu_2)}\big(\beta x_H + r\big), \quad \text{where } x_H = \frac{\lambda}{\mu_1}. \tag{6.2}$$

Moreover, in analogy to the characterisation of the proviral load as the infected fraction of all CD4$^+$ helper T-cells for the baseline model (3.1) (c.f. Equation (3.17)), we can also define for the multi-locus model (6.1) the infected fraction of all CD4$^+$ helper T-cells due only to strain ij as follows:

$$v_{ij} = \frac{u_{ij}+y_{ij}}{x + \sum_{i=1}^{m}\sum_{j=1}^{n}\big(u_{ij}+y_{ij}\big)}, \qquad \text{for } i = 1,\ldots,m \text{ and } j = 1,\ldots,n.$$

Next, as with the Tax/HBZ model (4.1), we define an expression ϕ_{epitope} (which we have previously called the effective CTL expansion rate; refer to Equation (4.3)), to describe CTL proliferation and lifespan. Respectively, for Tax-specific and HBZ-specific CTLs, we denote:

$$\phi_{\text{Tax}} = \frac{\nu}{\mu_4} \qquad \text{and} \qquad \phi_{\text{HBZ}} = \frac{\nu'}{\mu_5}. \tag{6.3}$$

6.2.3 The Special Case $n = 1$: Basic Analytical Insights Gained from the Multi-Locus HTLV-I System with Antigenic Variability in One Epitope

In this subsection, we briefly consider variability in only one of the epitopes. A thorough exploration of these interactions provides us with crucial insights that form a foundational basis for understanding multi-strain pathogens that will carry over more generally to the situation where antigenic variability is present at both epitopes.

Specifically, we examine two crucial aspects of within-host infection by HTLV-I, namely immunogenicity and inter-strain competition, in the special case when $n = 1$ (i.e. the so-called '$m \times 1$' model). Utilising a combination of analytical and computational techniques, our investigations reveal the following basic ideas (we refer the reader to Appendix A.4 for a detailed mathematical exploration of these three ideas):

1. **Chronic infection is only possible when the basic reproduction number, or fitness (i.e. R_{0,i_0}, where $i_0 \in \{1, \ldots, m\}$), of at least one viral strain is greater than unity.**

▷ Because we are considering the possibility of infection by multiple strains, the appropriate interpretation of chronic infection corresponds to the successful establishment of at least one strain (and not necessarily all strains concurrently). This interpretation and the results of our model are consistent with the notion of chronic infection in our previous models.

2. **In the absence of competition, the extent of immunogenicity of antigenic variants in an epitope is an important determinant of strain abundance, but not strain structure.**

▷ Related strains that have identical fitness can co-exist in linearly stable equilibrium with the host immune response, and the relative abundance or dominance of each strain is determined solely by the extent of immunogenicity of variant-specific responses.

3. **The introduction of inter-strain competition can result in different viral population structures due to the effects of competitive exclusion.**

▷ As soon as inter-strain competition is introduced, the ability of the less fit strain to persist depends strongly on the extent to which the dominant strain stimulates a variant-specific CTL response instead of a shared CTL response. There always exists a threshold degree of immunogenicity for which actively infected target cells from the dominant strain provide sufficient stimulation of the cross-reactive response to overwhelm and drive the weaker strain to extinction.

6.2.4 Characterisation of the Infection-Free Steady State in the General Multi-Locus Model

The following two theorems summarise the analytical results that we are able to establish regarding the existence and stability properties of steady state solutions of the multi-locus

model. The proofs of these results are analogous to those in the special case when $n = 1$ (see Appendix A.4 for details), and so we have chosen not to include them here.

Theorem 6.1 (Boundedness of Solution Trajectories for the Multi-Locus Model)**.** *Denote* $\vec{u} = \{u_{ij}\}_{i=1,\ldots,m}^{j=1,\ldots,n}$, $\vec{y} = \{y_{ij}\}_{i=1,\ldots,m}^{j=1,\ldots,n}$, $\vec{z} = \{z_i\}_{i=1,\ldots,m}$, *and* $\vec{s} = \{s_j\}_{j=1,\ldots,n}$. *Let* $\mu = \min\{\mu_1, \mu_2, \mu_3\}$, $\tilde{\nu} = \max\{\nu, \nu'\}$, *and* $\tilde{\mu} = \min\{\mu_4, \mu_5\}$, *and consider the closed, bounded region*

$$\Gamma_M := \left\{(x, \vec{u}, \vec{y}, \vec{z}, \vec{s}) \in \mathbb{R}_+^M : x \leq \frac{\lambda}{\mu_1},\ x + \sum_{i=1}^{m}\sum_{j=1}^{n}(u_{ij} + y_{ij}) \leq \frac{\lambda}{\mu - r},\right.$$
$$\left.\sum_{i=1}^{m} z_i + \sum_{j=1}^{n} s_j \leq \frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu - r)}\right\},$$

where $M = (2mn + m + n + 1)$. *Assume that* $r < \mu$. *Then, the set* Γ_M *defines a positively invariant region with respect to model* (6.1). *All solutions are bounded for* $t \geq 0$ *and eventually enter* Γ_M.

Theorem 6.2 (Global Dynamical Properties of the Infection-Free Equilibrium P_0 in the Multi-Locus Model)**.** *Let* $\vec{0}_N = (0, \ldots, 0) \in \mathbb{R}^N$ *denote the* N*-dimensional zero vector for some* $N \in \mathbb{N}$ *and let* $x_H = \lambda/\mu_1$. *The infection-free equilibrium for model* (6.1)*, given by* $P_0 = (x_H, \vec{0}_{mn}, \vec{0}_{mn}, \vec{0}_m, \vec{0}_n)$, *always exists. If* $0 < R_{0,ij} < 1$ *for all* $i = 1, \ldots, m$ *and* $j = 1, \ldots, n$, *then* P_0 *is the only equilibrium in* $\bar{\Gamma}_M$ *and it is globally asymptotically stable in* $\bar{\Gamma}_M$. *Conversely, if there exists* $(i_0, j_0) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$ *such that* $R_{0,i_0j_0} > 1$, *then* P_0 *is unstable. Specifically, let* $1 \leq k \leq m$ *denote the number of viral strains whose basic reproduction number is greater than unity. Then,* P_0 *is a saddle with* $\dim W_{loc}^s(P_0) = (2mn + m + n - k + 1)$ *and* $\dim W_{loc}^u(P_0) = k$, *where* $W_{loc}^s(P_0)$, $W_{loc}^u(P_0)$ *denote the respective local stable and unstable manifolds of* P_0.

6.2.5 A Remark on Analytic Tractability of the General Multi-Locus Model

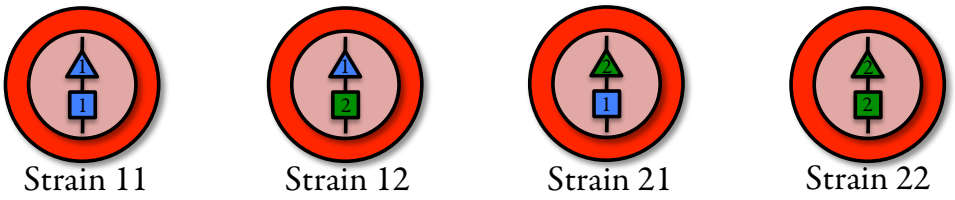
In developing the multi-locus HTLV-I model, we find that the trade-off between incorporating an increased level of biological complexity and retaining some degree of mathematical tractability becomes strikingly apparent. As a consequence of this, our investigation of the model relies heavily on numerical simulation and computational output with relatively fewer analytical results as compared with our previous models. Due to such a limitation, during our investigation of the model we have been careful in ensuring that the results presented in this chapter are the outcome of extensive simulation studies using parameter values drawn from the physiologically realistic ranges as outlined in Table 3.1. Importantly, we note that the various graphs and charts illustrated here are meant to be representative of the overall

dynamics of the multi-locus model (6.1) rather than being artefacts of a single, atypical simulation. Thus, our main focus in developing the multi-locus HTLV-I model is to build a consistent mathematical framework that can highlight a better conceptual understanding of the complicated interaction between host immunity and an antigenically variable pathogen. Our exposition of the results in this chapter is therefore somewhat different in style than that of previous chapters in the sense that the ideas presented here, although based on firm biological and experimental evidence, are slightly less rigorous from a purely mathematical perspective[2]. Inevitably, several of our interpretations are speculative, but nevertheless, we argue that our multi-locus modelling approach serves its intended purpose as it allows us to present and discuss biologically meaningful explanations of realistic aspects of HTLV-I infection which are both consistent with our multi-locus mathematical modelling framework and can be placed in the context of empirical evidence, leading to valuable insights from an immunological standpoint.

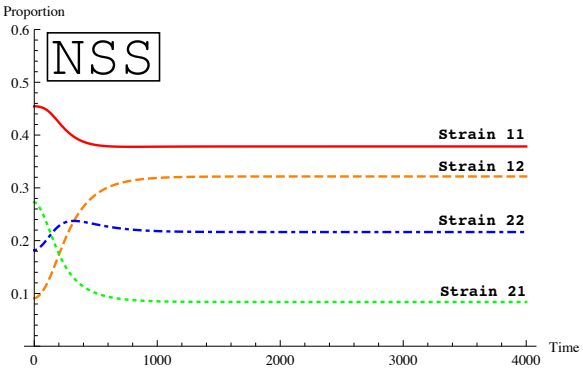
6.3 Co-existence of All Strains

We begin our examination of the multi-locus model (6.1) by considering the scenario in which there are no differences in fitness among the various virus strains, as illustrated in Figure 6.3. When all viral strains have identical fitness (shown schematically in Figure 6.3(a)), the successful establishment of one particular strain implies fixation of all species that are present at the onset of the infection, and the system naturally tends towards a stable chronic infection equilibrium where all strains co-exist at abundances dictated by the skew in immunogenicity (δ_i) of antigenic variants in the Tax epitope and the initial conditions of the system. In other words, there is *no strain structure (NSS)* present in the virus population. Figure 6.3(b) shows a typical time series graph. In the situation where all CTLs are functionally identical, the two CTL types achieve equilibrium at equal abundances when $\sum_{i=1}^{m} 1/\delta_i = 2m$, where m denotes the number of possible antigenic variants present in the Tax epitope. Figure 6.3(c) illustrates a heat map of CTL dominance for the situation with two variants in each of the two epitopes, where the degree of dominance of Tax-specific CTLs is determined simply by their relative abundance within the CTL pool. When both Tax-specific and HBZ-specific CTL types are equally abundant, or co-dominant, the graph demonstrates numerical agreement at the theoretically derived boundary curve, shown by the black, dashed line (see Appendix A.5.1 for a derivation of the general result).

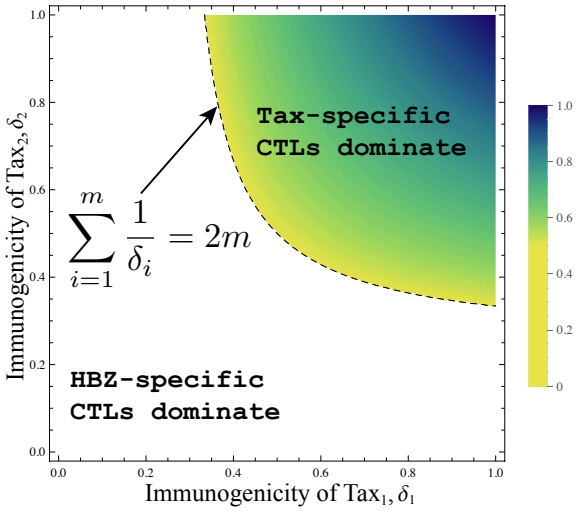
[2]It should be noted that several authors have developed some analytical insights in other systems that exhibit a multi-locus model structure (Calvez et al., 2005; Recker and Gupta, 2005; Blyuss and Gupta, 2009); however, their work is not applicable to our model.



(a) All strains have identical fitness.



(b) No strain structure (NSS).



(c) CTL Dominance.

Figure 6.3: With respect to the multi-locus model (6.1), in the absence of inter-strain fitness discrepancies, we observe the co-existence of all viral strains and no strain structure (NSS). (a) Assume that no fitness differences exist among the strains. (b) Typical time evolution graph showing convergence to a chronic infection steady state in which all possible strains co-exist. Here, the abundance of each viral sub-population is determined by the skew in immunogenicity δ_i of Tax variant i as well as the initial conditions of the system. (c) Heat map showing the degree of dominance of Tax-specific CTLs, which is represented by their relative abundance within the pool of all CTLs. The two types of epitope-specific CTL responses maintained equilibrium at equal frequencies when $\sum_{i=1}^{m} 1/\delta_i = 2m$, where m is the number of antigenic variants in the Tax epitope. In the above figures, $R_{0,ij} = 2.56$ for all strains. Values for the various parameters describing each strain in model (6.1) are selected from biologically relevant ranges as in Table 3.1.

However, more detailed calculations reveal that degeneracy can arise when there is no inter-strain variability in the multi-locus model (6.1). Specifically, we have found that in the absence of fitness differences among the various strains, there is an infinite line of equilibria in which all possible strains co-exist. By contrast, the introduction of inter-strain differences in fitness precludes the existence of such interior equilibria, meaning that at least one subpopulation of proviral cells is eventually outcompeted by the other strains and eliminated by host immunity. The precise statement and proof of this result can be found in Appendix A.5.2.

6.4 The Simplest Example of a Multi-Locus System: 2 Variants in Tax, 2 Variants in HBZ

To illustrate the main features of our multi-locus HTLV-I model (6.1), we examine the situation in which there are two possible forms for the Tax protein (say, Tax_1 and Tax_2), and two possible forms for the HBZ protein (say, HBZ_1 and HBZ_2). The result is four distinct virus variants: $\text{Tax}_1/\text{HBZ}_1$, $\text{Tax}_1/\text{HBZ}_2$, $\text{Tax}_2/\text{HBZ}_1$, and $\text{Tax}_2/\text{HBZ}_2$. We will denote these as variants 11, 12, 21, and 22, respectively, and we refer to this model as the 2×2 model.

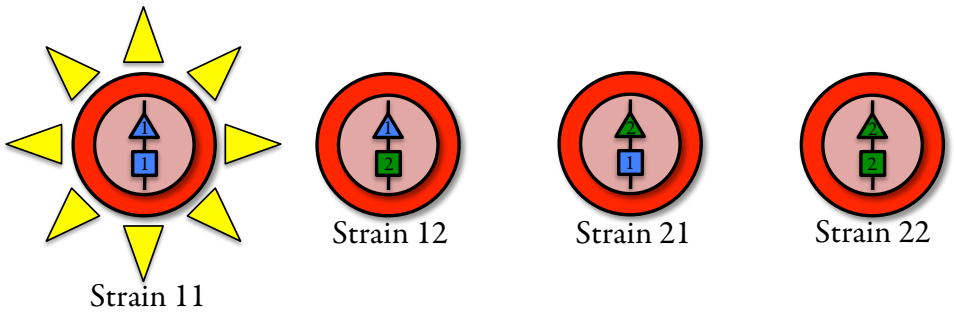
The existence of the four distinct virus variants as described above is the simplest way in which we can represent our multiple epitope model with antigenic variability in both Tax and HBZ. By exploring how these virus variants interact with each other and with antigen-specific host immunity, we hope to build a conceptual understanding of the multi-locus interactions in HTLV-I infection. The underlying principles are generalisable to the case when any number of distinct antigenic variants exist at each of the two epitopes.

6.4.1 Differences in Inter-Strain Fitness Create Patterns of Viral Strain Structure

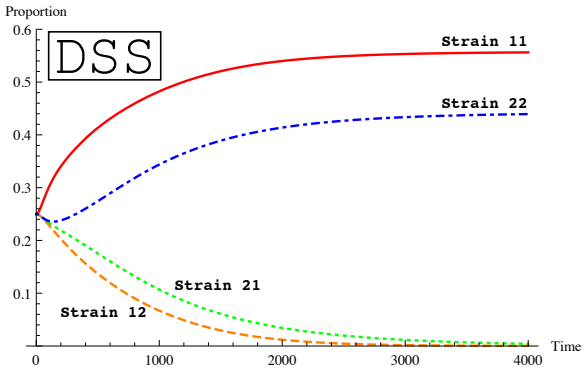
In the previous section, we have discussed the outcomes of the multi-locus model in the case when there are no differences in fitness among the viral strains. However, experimental measurements of T-cell kinetics have provided strong evidence that viral sub-populations can exhibit variability in their individual levels of fitness; for example, it has been shown that mutations or nucleotide substitutions in the *tax* gene can affect Tax-mediated transactivation activity (which is a measure of viral fitness) through the HTLV-I long terminal repeat (LTR) (Furukawa et al., 2000; Renjifo et al., 1995; Semmes and Jeang, 1992). Indeed, such inter-strain variability is a direct consequence of differences in the way that the combination of distinct protein variants at each epitope (e.g. Tax_1 versus Tax_2 at the Tax epitope and HBZ_1 versus HBZ_2 at the HBZ epitope) contributes to the proper functionality of cellular processes of the viral strain possessing that particular set of antigenic variants and therefore determine its basic reproductive ratio.

In the more general situation where there exists heterogeneity in the basic reproductive ratios among distinct viral strains, we have proved that the co-existence of all viral sub-populations is not possible and the HTLV-I-infected cell pool will instead self-organise into ordered sets in which at least one strain is eliminated as the system equilibriates (refer to Appendix A.5.2), an outcome that arises as a consequence of the competitive exclusion principle (Hardin, 1960). In the following, we assume that in the 2×2 model, only strain

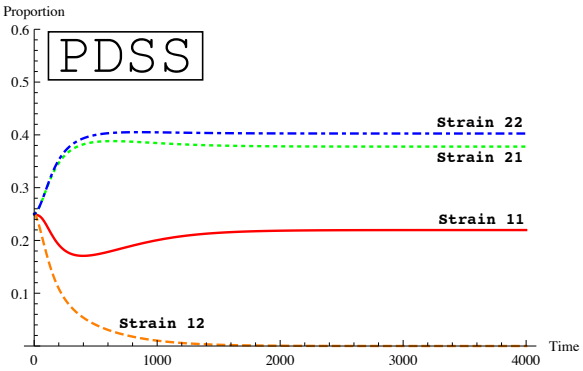
11 has a fitness advantage compared to the remaining strains (see Figure 6.4(a) for an illustration). Our simulation results here appear to indicate that there emerge two qualitatively distinct patterns of strain structure that depend on the relative degrees of immunogenicities (δ_1 and δ_2) towards each one of the two Tax epitopes: *discrete strain structure (DSS)*, and what we will refer to as *partial discrete strain structure (PDSS)*. The time evolution graphs for strain abundances from the start of infection to the establishment and maintenance of each of the two types of viral strain structures are shown in Figures 6.4(b)–(c).



(a) Strains have different fitness.



(b) Discrete, non-overlapping strain structure.



(c) Partial discrete, overlapping strain structure.

Figure 6.4: Heterogeneity in strain fitness creates structure in the viral population in the multi-locus model (6.1). (a) To illustrate the main principles underlying the role and outcome of inter-strain differences, we assume for the meantime that only one of the virus strains, specifically strain 11, has a fitness advantage over the other strains. (b)–(c) Representative time evolution plots for the solution trajectories of model (6.1). Extensive numerical investigations of the model show the emergence of two qualitatively distinct types of convergent strain structures which arise as a consequence of inter-strain differences in fitness: (b) discrete, non-overlapping strain structure (DSS), and (c) partial discrete, overlapping strain structure (PDSS). The type of strain structure achieved, as well as the abundances of each individual strain, are determined by the relative differences in immunogenicities towards each of the Tax protein variants. The graphs show typical time evolution plots of the virus population in each of the two regions.

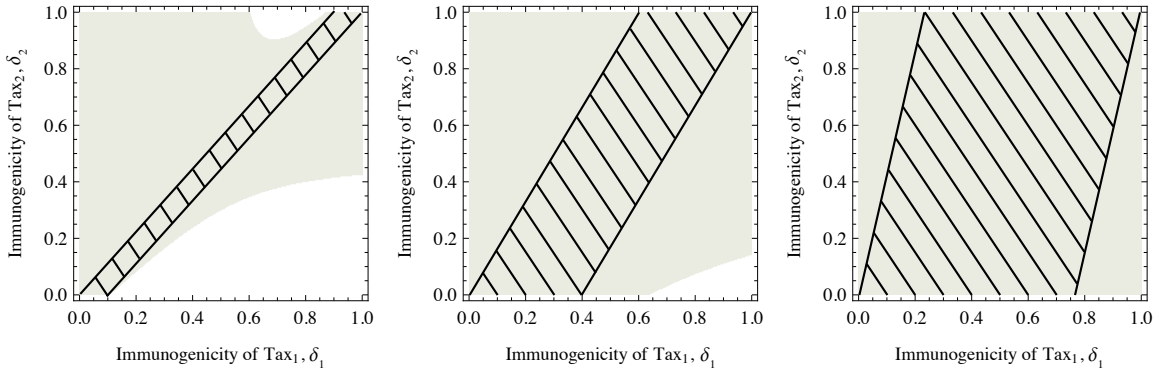
DSS, sometimes referred to as discrete, non-overlapping strain structure, is characterised by the stable co-existence at equilibrium of a set of strains that do not share antigenic

variants at any of the epitopes. For example, the non-overlapping pair of strains, namely, strains 11 and 22, exhibit DSS (see Figure 6.4(b)). Alternatively, PDSS describes a proper subset of all strain types that express overlapping variants in at least one of the epitopes. In other words, PDSS consists of a discrete set of strains along with a number of other strains that share a protein variant at either the Tax epitope or HBZ epitope. For instance, the overlapping set of strains, namely, strains 11, 21, and 22, represents an example of PDSS (see Figure 6.4(c)). We will refer to the convergence to a particular organised pattern of co-existing strains during the chronic phase of infection, whether DSS or PDSS, as establishing a *stable strain structure*.

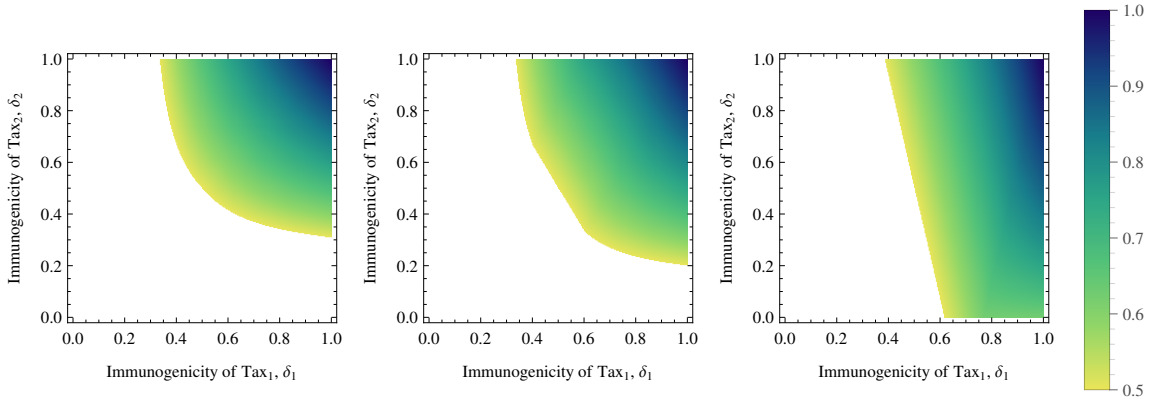
6.4.2 Discrete Strain Structure, Strain Dominance, and Stratification of the HTLV-I-Specific CTL Response

The consideration of HTLV-I as a multi-locus system highlights two complementary aspects of the host-virus interaction. On the one hand, the simultaneous expression of multiple viral epitopes means that a single strain can be recognised by multiple distinct CTL responses. On the other hand, because distinct strains may share common protein variants at one epitope, a single variant-specific CTL response is also able to suppress multiple different strains (we refer the reader back to the schematic illustration in Figure 6.2). As we have seen, these two aspects not only contribute to the establishment and maintenance of a certain type of strain structure (in the presence of inter-strain differences in fitness), but also influence the composition of the virus population. In other words, the combination of immune-mediated pressures exerted by the various CTL types on each individual virus strain can create wide variation in the relative abundances of each strain, even whilst maintaining a specific viral population structure (i.e. DSS or PDSS). Moreover, an important aspect of our modelling approach is that we are able to characterise the immune profile of an infected host. Since HTLV-I-infected cells are under positive selection by a host cellular immune response which is highly specific to target antigens, the existence of protein variants at distinct antigenic loci in the virus population therefore implies a corresponding measure of diversity in anti-HTLV-I CTL responses.

Figure 6.5 captures a series of representative plots illustrating one of the main results from our multi-locus HTLV-I model, namely that the organisation and abundance of both virus and CTLs during chronic infection are strongly influenced by two primary factors: (i) the extent of immunogenicity of protein variants at the immunodominant Tax epitope (i.e. δ_1 and δ_2), and (ii) the degree of heterogeneity in the level of fitness of individual strains (i.e. $R_{0,ij}$ values). The results shown in Figure 6.5 are the outcome of a typical simulation as described below.



(a) Strain Structure and Fittest Strain Dominance.



(b) Tax-Specific CTL Dominance.

Figure 6.5: Numerical investigation of the multi-locus model (6.1) indicates that the degree of heterogeneity in the fitness of individual strains (i.e. $R_{0,ij}$ values) and the extent of immunogenicity towards protein variants in the Tax epitope (i.e. δ_i) play a significant role in determining the structure of both virus and CTL populations. Here, we have illustrated the case when $m = n = 2$, i.e. when there exist two protein variants at each epitope, which we call the 2×2 model. (a) In conjunction with heterogeneity in the fitness levels among the virus strains, the immunogenicities of the various antigenic variants at the Tax epitope determine the type of organised strain structure that the system will adopt: DSS, or PDSS. DSS is shown by cross-hatching, while the absence of cross-hatching designates PDSS. The shaded areas indicate regions for which the strain with the highest fitness, in terms of the basic reproductive ratio, dominates the virus population (defined as the strain that persists at the highest prevalence). The parameters δ_i represent the respective immunogenicities towards variant i at the Tax epitope. Our results demonstrate that a large fitness discrepancy between the most advantageous strain and the remaining strains increases the region for which DSS is stable, i.e. the system converges to an equilibrium describing DSS. The most advantageous strain is not necessarily dominant in the PDSS region when the inter-strain fitness discrepancy is sufficiently low. (b) Immune selection on an antigenically variable virus population creates organised CTL structure. The coloured regions indicate parameter regimes for δ_i where Tax-specific immune responses (all sub-types) dominate the CTL population at equilibrium, i.e. comprises more than half of the total CTL population. The intensity of Tax-specific CTL dominance is represented by a colour gradient as indicated. In the above figures: Left panel, $R_{0,11} = 2.71$; Centre panel, $R_{0,11} = 3.46$; Right panel, $R_{0,11} = 6.61$. For all other strains, $R_{0,ij} = 2.56$, where $ij \neq 11$. Parameters used to simulate the model have been drawn from biologically realistic ranges as discussed in Table 3.1.

In Figure 6.5(a), we illustrate the regions corresponding to a particular type of 'stable strain structure'. We observe that when the immunogenicities of the Tax variants are similar with respect to the differences in inter-strain fitness, the virus population adopts a pattern of DSS (cross-hatched regions in Figure 6.5(a); the time series graph corresponding to any specific point in the DSS region is qualitatively similar to the one shown in Figure 6.4(b)). Alternatively, when the Tax variants have sufficiently skewed immunogenicities, the virus population instead converges to PDSS (the regions that are not cross-hatched in Figure 6.5(a); a point lying in the PDSS region has a corresponding time series graph which is qualitatively similar to the one in Figure 6.4(c)). In this latter scenario, the persistence of a specific combination of overlapping viral strains depends on whether strong competition is favoured at the Tax locus or the HBZ locus, resulting in competitive exclusion of only those strains that share either a common Tax variant or HBZ variant with the most fit strain[3]. Meanwhile, the shaded areas indicate parameter choices for δ_1 and δ_2 for which the strain with the greatest fitness, with respect to the basic reproductive ratio, dominates the virus population (we consider a single strain as dominant if it persists at the highest prevalence amongst all the strains). Our numerical investigations indicate, at least for the selected parameter values, that DSS can persist for a wide range of differences between the immunogenicities of the two Tax variants and, moreover, the range for which such a set of non-overlapping strains exists at stable equilibrium with host immunity increases as the discrepancies in strain fitness widen (compare the sizes of the cross-hatched regions among the three panels in Figure 6.5(a)).

Figure 6.5(b) illustrates the results of solving model (6.1) numerically until it has reached a stable steady state and examining the structure of Tax-specific CTLs with respect to the immunogenicities, δ_1 and δ_2, of each of the two Tax variants for varying degrees of heterogeneity in fitness among the strains. The coloured regions in the heat maps in Figure 6.5(b) designate parameter regimes for δ_1 and δ_2 in which Tax-specific cellular immune responses (targeting any of the protein variants at the Tax locus) dominate the CTL population, in terms of frequency, during chronic infection and represent a gradient (from yellow to green to dark blue), where the lightest shade of yellow along the edge of each region reflects the situation where the pool of Tax-specific CTLs weakly dominate that of HBZ-specific CTLs and the darker shades of blue are indicative of a response that is predominantly comprised of CTLs that recognise only protein variants at the Tax epitope. For instance, in the extreme case where both δ_1 and δ_2 are equal to 1 (i.e. both Tax protein variants are fully

[3]We have not proved this explicitly for the multi-locus model. However, we briefly touched upon how inter-strain competition between any two strains that share a common antigenic target can lead to competitive exclusion of the lesser fit strain during our investigation of the '$m \times 1$' model as described in Subsection 6.2.3. We speculate that the same basic principles apply to the multi-locus model (6.1), and numerical investigation of the model appears to corroborate such a proposition.

immunodominant), the host immune cell population is made up entirely of Tax-specific CTLs.

We highlight three details of interest based on our observations of the simulation results of the 2×2 model presented in Figure 6.5.

1. In the region where DSS occurs (see Figure 6.5(a), cross-hatched areas), the fittest strain is always dominant (Figure 6.5(a), shaded areas). This observation points to DSS patterning as a route by which HTLV-I can promote viral homogeneity. In addition, as the discrepancy between strain fitnesses is increased, not only does the region for which DSS arises become larger, so does the region for which the most fit strain dominates the virus population (compare Figure 6.5(a), left to right panels).

2. In the regions where PDSS persists, the model results demonstrate the proof of principle that the most advantageous strain need not necessarily be dominant and, in fact, other strains that have no fitness advantage can exist at a higher abundance within the virus population. This observation, despite appearing somewhat counter-intuitive at first, should not be entirely unexpected. The biological reasoning is as follows: The fitness of a particular strain is determined by the functionality of the specific combination of protein variants at each epitope, and a virus strain that has a higher level of fitness possesses an advantage in its ability to productively propagate its lineage to a subsequent generation. However, the action of variant-specific CTL responses is impacted greatly by epitope immunogenicity. A virus strain displaying epitopes that are highly immunogenic elicits a vigorous host immune response and can therefore be suppressed to low levels, regardless of whether or not it has a fitness advantage.

3. Our numerical investigation of the CTL population structure appears to indicate that as the discrepancy in strain fitness increases, Tax-specific cellular immune responses are able to dominate the CTL population provided only that the immunogenicity δ_1 of the Tax protein variant expressed by the most advantageous strain is sufficiently large, regardless of the degree of immunogenicity δ_2 of the other Tax protein variant.

Remark 6.2. With respect to the second detail of interest raised above (i.e. Point 2), the results from our 2×2 model help illustrate this point. We refer the reader to Figures 6.6 and 6.7, which represent, in the form of pie charts, the proportions of each strain for five evenly spaced, selected sample values of δ_1 and δ_2 from the three panels in Figure 6.5(a) (Figure 6.6 is simply a chart legend explaining how to interpret the pie charts in Figure 6.7). We refer to this style of visualisation as a '*concentric pie chart representation*'. In Figure 6.7, we notice that when Tax variant 1 is sufficiently more immunogenic than

Tax variant 2, strain 11 (designated as a light blue colour in the pie chart representation), despite being the most fit strain, is the least abundant one, at least in the case when inter-strain fitness discrepancies are sufficiently low (outermost pie charts). This observation can be straight-forwardly explained: strain 11 expresses Tax variant 1 and HBZ variant 1. A strongly immunogenic Tax variant 1 (i.e. large value of δ_1) elicits a vigorous Tax-specific CTL response targeting all strains expressing Tax variant 1, including Strain 11. Meanwhile, a weakly immunogenic Tax variant 2 (i.e. small value of δ_2) means that all strains expressing Tax variant 2 elicit host immune responses that target respective variants in their HBZ epitopes. Some of these HBZ-specific CTL responses, namely the one specific to HBZ variant 1, can also target strain 11. We hypothesise, based on our model observations, that compounded host immunity acting against strain 11 leads to its suppression, whereas other strains that are not subject to intense immune-mediated selection are given freedom to propagate.

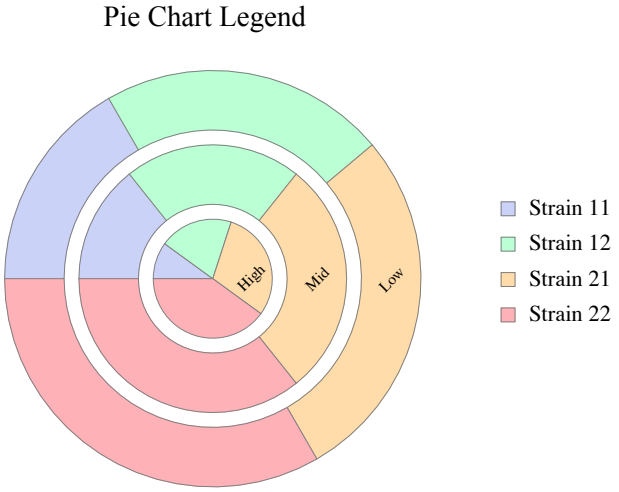


Figure 6.6: Pie chart legend for the composition of the virus population in terms of relative abundances of each individual viral strain for different degrees of inter-strain fitness discrepancies. This schematic legend illustrates how we visualise the structure of the virus population at equilibrium according to the multi-locus model (6.1) for a fixed combination of immunogenicities of the two Tax variants, i.e. a fixed choice of δ_1 and δ_2. Each circular plot, as shown in the schematic figure above, consists of three separate, concentric pie charts detailing the respective virus population structures corresponding to low (outermost pie chart), intermediate (middle pie chart), and high (innermost pie chart) differences in fitness between the most advantageous strain and the remaining strains. The four different colours represent the individual strains as indicated. We will call this style of visualising the model results a 'concentric pie chart representation'.

Our numerical simulations demonstrate that the phenomena described above can arise when all strains have similar levels of fitness or, in other words, when the inter-strain

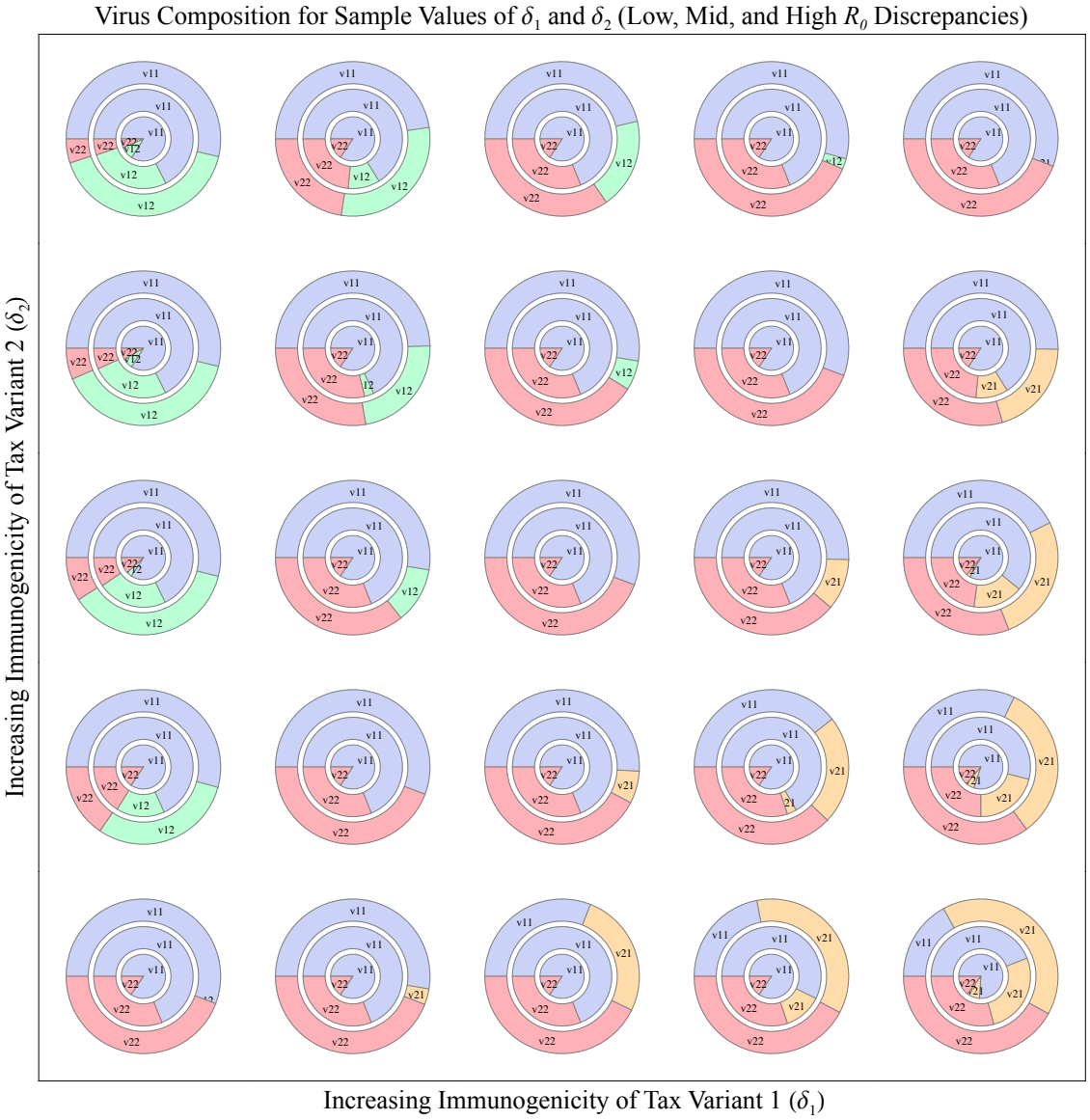


Figure 6.7: Examining virus composition for varying degrees of inter-strain fitness discrepancies. Here, we have selected five uniformly spaced immunogenicity values for both δ_1 and δ_2 within the interval $(0,1)$ (for a total of $5\times 5 = 25$ combinations). For each δ_1, δ_2 combination, we consider three degrees of inter-strain fitness discrepancies: low, intermediate, and high, which are determined by the differences in fitness among the various viral strains, e.g. a low inter-strain fitness discrepancy means that the least fit strain and the most fit strain have similar basic reproduction numbers. With fixed values of δ_1, δ_2, and the degree of inter-strain discrepancy (i.e. low, intermediate, or high), we then solve the 2×2 model numerically, and plot the relative abundances of each individual strain as a single pie chart element in the concentric pie chart representation as described in the schematic in Figure 6.6. Thus, the figure shows the outcomes of $25\times 3 = 75$ simulations. **(Caption continued on following page.)**

differences in fitness are sufficiently low (refer to Figure 6.7, outermost pie charts). However, our results suggest that provided there is a sufficiently large discrepancy in fitness between

Figure 6.7: **(Continued caption from preceding page.)** The columns and rows show the strain structures corresponding to a fixed immununogenicity of Tax variant 1 (i.e. δ_1), and of Tax variant 2 (i.e. δ_2), respectively. Meanwhile, the outermost, middle, and innermost pie charts show the resulting strain structures associated with low, intermediate, and high differences in fitness among the strains (for a specified combination of δ_1 and δ_2). To put the figure into perspective, we note that each of the individual pie charts above is simply detailing the composition of the virus at a particular point in the strain structure graphs plotted in Figure 6.5(a) for low, intermediate, and high inter-strain fitness discrepancies (corresponding to the left, centre, and right panels of Figure 6.5(a), respectively). The basic reproduction numbers for the various strains are uniquely determined by their individual rates of spontaneous viral expression, as represented by the parameter τ_{ij}. The quantity, v_{ij}, refers to the fraction of target cells infected by strain ij, and is defined as follows: $v_{ij} = \big(u_{ij} + y_{ij}\big)/\Big(x + \sum_{i=1}^{m}\sum_{j=1}^{n}(u_{ij} + y_{ij})\Big)$. In the above figure, the v_{ij} for the various strains in each pie chart element are shown as relative proportions. The values of the basic reproduction numbers for the fittest strain 11 are as follows: outermost pie charts, $R_{0,11} = 2.71$; middle pie charts, $R_{0,11} = 3.46$; innermost pie charts, $R_{0,11} = 6.61$. For all other strains, $R_{0,ij} = 2.56$, where $ij \neq 11$. Moreover, δ_1 and δ_2 take the specific values 0.05, 0.275, 0.5, 0.725, and 0.95. We observe that the composition of the virus population becomes more homogeneous as inter-strain fitness increases (compare outermost versus innermost pie charts).

the most fit strain and all other strains, the strain with the highest fitness is likely to always dominate the viral population (refer to Figure 6.7, innermost pie charts).

6.4.3 Differences in Effective Expansion Rates Between CTLs Determine Strain and CTL Regions

The effective expansion rate of each type of anti-viral immune response is a measure of its proliferative capacity and lifespan, and is governed by both genetic factors as well as the effects of inter-cellular signalling. For instance, in terms of CTL proliferation, the activation of a pool of virus-specific CTLs by its cognate antigen leads to rapid clonal expansion of that particular CTL type (Parham, 2005). Alternatively, while mechanisms such as activation-induced programmed cell death reduce CTL lifespan, other inter-cellular processes, including the engagement of inhibitory killer-cell immunoglobulin-like receptors (KIRs) on natural killer cells with virus-specific CTLs, have recently been suggested to prolong the lifespan of CTLs (Seich al Basatena et al., 2011; Ugolini et al., 2001). Differences between CTLs in their lifespans introduce heterogeneity in the effector cell pool and can thus shift the dynamical properties of the virus population. An increased CTL lifespan (and hence an increased effective expansion rate) means that such effector cells spend more time suppressing virus-expressing target cells and could therefore have a substantial effect on the long-term outcome of infection.

Using our multi-locus model, we explore the way in which differences in the proliferation and lifespan ratios (collectively grouped together as the effective expansion rate) of Tax-specific and HBZ-specific CTLs, represented respectively by ϕ_{Tax} and ϕ_{HBZ}, structure the

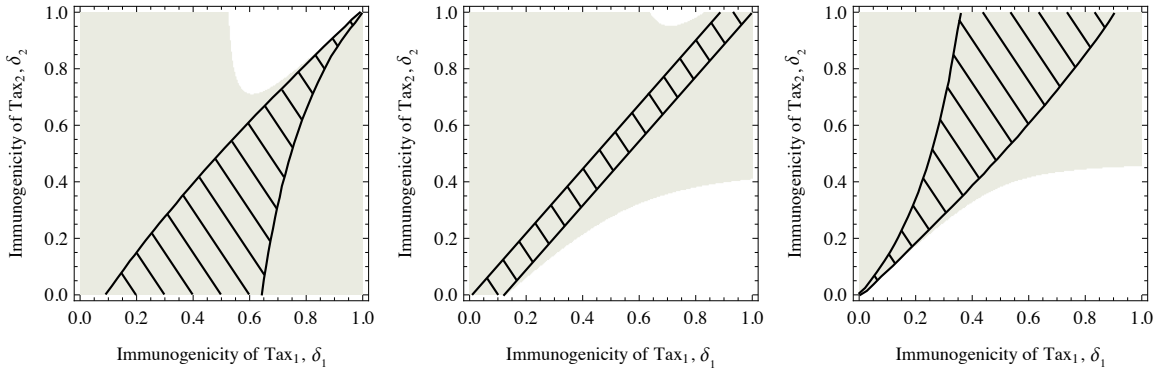
virus and CTL populations. The results of a typical simulation are illustrated in Figure 6.8. We observe that a change in the shape of the regions for which DSS and PDSS are stable depends on the ratios of the effective expansion rates between the two types of epitope-specific CTL responses (Figure 6.8(a)). The dominance of Tax-specific CTLs, in terms of relative CTL abundance, is also strongly influenced by the relationship between the different CTL types (Figure 6.8(b)). Indeed, if the ratio of effective expansion rates is skewed towards HBZ-specific CTLs (i.e. $\phi_{\text{Tax}}/\phi_{\text{HBZ}} < 1$), then the DSS region emerges for higher immunogenicity values of δ_1 (the Tax variant presented by the fittest strain 11), and lower immunogenicity values of δ_2 (the Tax variant not presented by the fittest strain) and, moreover, the CTL population is largely dominated by CTL effectors targeting HBZ, even for high degrees of Tax variant immunogenicities (i.e. when δ_1 and δ_2 are close to 1).

Meanwhile, if the skew of CTL effective expansion rates is biased towards Tax-specific CTLs (i.e. $\phi_{\text{Tax}}/\phi_{\text{HBZ}} > 1$), then the converse holds; that is, DSS is favoured for lower δ_1 and higher δ_2 and, additionally, the dominant CTL type consists of those targeting Tax, even when the immunogenicity values of Tax variants are low (i.e. when δ_1 and δ_2 are close to 0; see Figure 6.8, right column). In each of the two situations described above, our simulations demonstrate that a greater skew between ϕ_{Tax} and ϕ_{HBZ} results in a larger curved region of DSS.

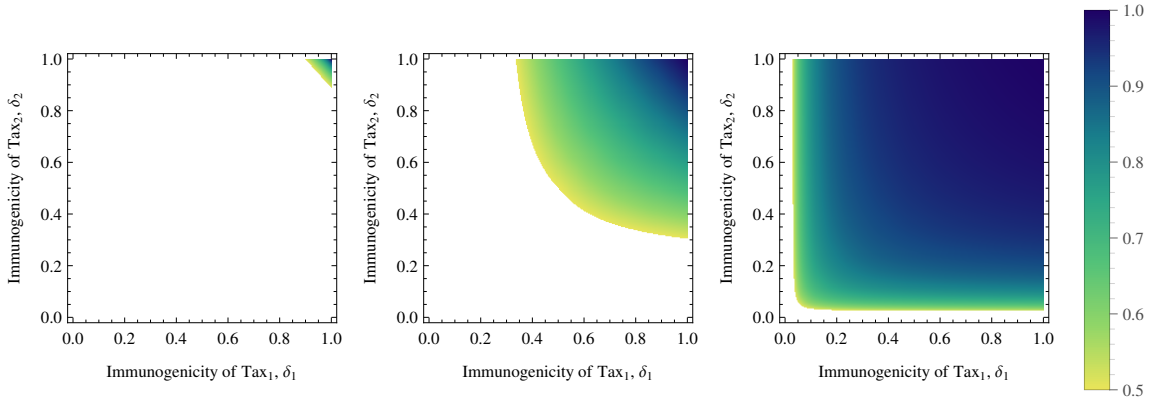
Lastly, when both CTL types display the same effective expansion rates (i.e. $\phi_{\text{Tax}}/\phi_{\text{HBZ}} = 1$), the DSS and PDSS regions, as well as the corresponding CTL dominance heat plot, obey the same rules as we have examined previously with respect to inter-strain differences in fitness (in other words, the plots resemble those seen in Figure 6.5).

Remark 6.3. If the effective expansion rates of both CTL types are very low (e.g. CTL lifespans are $\sim$1 day), then the shape of the region where Tax-specific CTLs are dominant is similar to what is seen with fitness discrepancies, i.e. the right-hand side of the δ_1, δ_2-plot is shaded instead of just the top-right quadrant. In particular, Tax-specific CTLs will dominate as long as the Tax variant 1 is sufficiently immunogenic, and the immunogenicity of Tax variant 2 can be lower. We speculate that this occurs because low effective CTL expansion rates imply fewer circulating CTLs (which interact with virus-infected cells), and suggest that differences among virus strains play a larger role (e.g. the impact of inter-strain competition).

Remark 6.4. In addition to the impact of CTL expansion on virus and CTL population structure, our numerical simulations indicate that there also exists a positive correlation between effective CTL expansion rate and the ratio of CTLs to activated proviral cells (results not shown). We found that this positive correlation holds in both the DSS and



(a) Strain Structure and Fittest Strain Dominance.



(b) Tax-Specific CTL Dominance.

Figure 6.8: Numerical investigation of the multi-locus model (6.1) show that relative differences in the effective expansion rates of immune effector cells and the skew in immunogenicity towards protein variants in the Tax epitope (i.e. δ_i) determine the organisation of the virus and CTL populations in the presence of inter-strain differences in heterogeneity. Here, we have illustrated the output from a typical simulation for the 2×2 model using physiologically realistic parameters as discussed in Table 3.1. The effective expansion rate for each class of CTLs is a ratio of its rate of proliferation and rate of natural decay. We consider the relative difference between the effective expansion rates of Tax-specific to HBZ-specific CTLs, which is given by the ratio $(\phi_{\text{Tax}}/\phi_{\text{HBZ}})$. (a) The areas for which DSS is stable are shown by cross-hatching, while the shaded areas indicate regions for which the strain with the highest fitness, in terms of the basic reproductive ratio, dominates the virus population (defined as the strain that persists at the highest prevalence). (b) Immune selection on an antigenically variable virus population creates organised CTL structure. The coloured regions indicate parameter regimes for δ_i where Tax-specific immune responses (all sub-types) dominate the CTL population at equilibrium, i.e. comprises more than half of the total CTL population. The intensity of Tax-specific CTL dominance is represented by a colour gradient as indicated. In agreement with current immunological understanding of cellular immunity, we observed that the degree of anti-Tax CTL dominance is positively correlated with the immunogenicities, δ_i, of antigenic variants in the Tax epitope and, furthermore, is greater as the skew of effective CTL expansion is biased towards Tax-specific CTLs. Left column: $\phi_{\text{Tax}}/\phi_{\text{HBZ}} < 1$. Centre column: $\phi_{\text{Tax}}/\phi_{\text{HBZ}} = 1$. Right column: $\phi_{\text{Tax}}/\phi_{\text{HBZ}} > 1$. Specifically, the values for the ratio $\phi_{\text{Tax}}/\phi_{\text{HBZ}}$ in the left, centre, and right columns are, respectively, 1/18, 1, and 18. In the above figures, $R_{0,11} = 2.71$, $R_{0,\text{others}} = 2.56$.

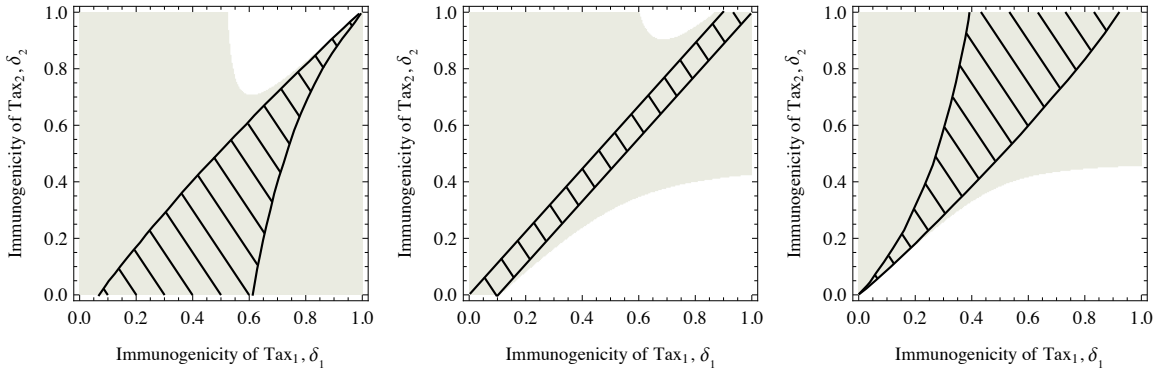
PDSS regions, and is also robust to both small and large changes in inter-strain fitness discrepancies. Although we have not proved this relationship analytically, our observation echoes an analogous result from the basic two-epitope Tax/HBZ model (4.1) (in particular, refer to Subsection 4.3.1).

6.4.4 Heterogeneity in the Efficiencies of Epitope-Specific CTL Responses Shapes the Regions for Ordered Strain and CTL Structure

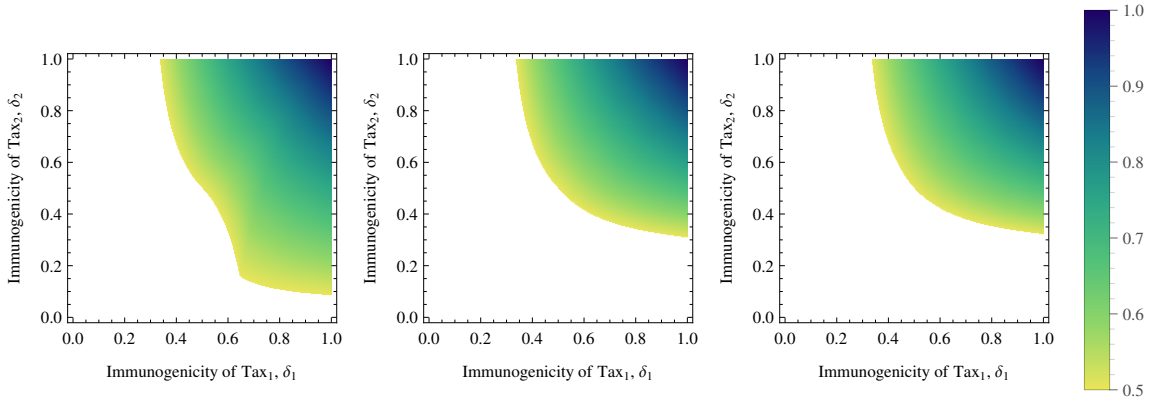
One specific prediction of the multi-locus model is that differences among individual strains induce heterogeneity in the virus population and allow HTLV-I to self-organise into structured patterns consisting of multiple strains (not all possible strains co-exist, but also not just one; see Subsection 6.4.1). Considering the significance of host immunity in the dynamics of HTLV-I infection, we raised the question: How does the strength, or efficiency, of epitope-specific CTLs influence the patterns of structure and dominance observed in viral strains and host immune responses?

Figure 6.9 shows a representative simulation illustrating how the regions where each of the two types of strain structures occur can be shaped by host immunity, assuming that strain 11 has a slight fitness advantage over the other strains. A noteworthy observation is that the change in shape of the regions is not dependent on the strengths of each epitope-specific CTL response, but rather on the relative differences in strength between the CTL responses that target Tax versus those that target HBZ. Indeed, when both CTL types have equal lytic rates so that $\gamma = \gamma'$, the shape of the DSS and PDSS regions are the same as governed by inter-strain fitness discrepancies, with the area of the DSS region inversely proportional to CTL strength (see Figure 6.9, centre column, and compare with Figure 6.5(a)). Namely, a more efficient rate of CTL-mediated lysis is associated with a decrease in the area for which the DSS is stable. However, if the CTL strength ratio is skewed towards HBZ-specific CTLs (i.e. if $\gamma/\gamma' < 1$), then the shape of the DSS region changes to favour higher immunigenicity values of Tax_1, the variant possessed by the advantageous strain 11, and lower immunogenicity values of Tax_2, the variant which is not possessed by the advantageous strain (see Figure 6.9, left column). Conversely, if CTL-mediated selection (either Tax-specific or HBZ-specific) is more effective against the Tax epitope rather than the HBZ epitope, in which case $\gamma/\gamma' > 1$, then DSS is favoured for lower immunogenicity values of Tax_1 and higher immunogenicity values of Tax_2 (see Figure 6.9, right column).

Remark 6.5. As with the case of comparing the effective CTL expansion rates of the two epitope-specific CTL types, our numerical investigations demonstrate that a greater skew in CTL efficiencies, i.e. γ/γ', in either direction results in a larger curved region for which DSS occurs. Comparison of Figure 6.8(a) with Figure 6.9(a) implies that CTL efficiency and



(a) Strain Structure and Fittest Strain Dominance.



(b) Tax-Specific CTL Dominance.

Figure 6.9: Numerical investigation of the multi-locus model (6.1) demonstrating that heterogeneity in the efficiencies of CTL responses can change the shape of the regions of organised strain structure and CTL dominance, and is dependent primarily on the strength ratio between Tax-specific and HBZ-specific CTLs (γ/γ'). Here, we have illustrated the outcome of a representative simulation of the 2×2 model, which has been parameterised using physiologically realistic values taken from Table 3.1. (a) The cross-hatched regions indicate DSS and the shaded regions represent dominance of the fittest strain (in terms of its persistence at the highest prevalence among all strains). Comparison to Figure 6.8(a) suggests that the ratios of both CTL efficiencies and effective CTL expansion rates between the two epitope-specific CTL types affect the establishment of strain structure, either DSS or PDSS, in similar ways. (b) CTL dominance and structure can also be influenced by the skew in efficiencies between the CTL types. The heat maps indicate parameter regimes for δ_1 and δ_2 where Tax-specific immune responses (all sub-types) dominate the CTL population at equilibrium, i.e. comprises more than half of the total CTL population. The intensity of Tax-specific CTL dominance is represented by a colour gradient as indicated. Left column: $\gamma/\gamma' < 1$. Centre column: $\gamma/\gamma' = 1$. Right column: $\gamma/\gamma' > 1$. Specifically, the values for the ratio γ/γ' in the left, centre, and right columns are, respectively, 1/16, 1 and 16. $R_{0,11} = 2.71$, $R_{0,\text{others}} = 2.56$.

effective CTL expansion (including CTL lifespan) may play a similar role in determining the strain structure of HTLV-I.

6.4.5 Proviral Load, Viral Composition, and Strain Dominance Are Differentially Influenced by CTLs Depending on Type of Strain Structure

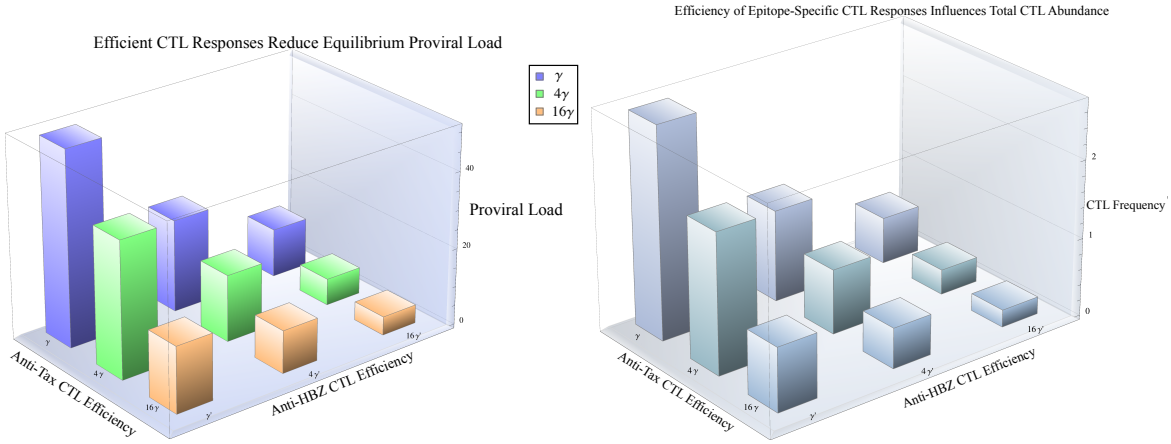
The efficiency of the HTLV-I-specific cellular immune response is a direct measure of its strength and has been shown to be a significant factor in effectively suppressing viral spread (Asquith et al., 2005a). We therefore explored in more detail the role that CTL strength plays in determining viral structure, total abundance, as well as the prevalence of the fittest strain in the context of the multi-locus HTLV-I model (6.1).

Figures 6.10, 6.11 and 6.12 plot typical outcomes of numerical investigation of the above factors for the 2×2 model. These results show that differing strengths of each type of epitope-specific CTL response can affect the magnitude and composition of the virus population in both the DSS and PDSS regions. We first observe that in both the DSS and PDSS settings, more efficient CTL responses of either specificity, either Tax-specific or HBZ-specific, are associated with a reduced proviral load and, as the pool of CTLs is maintained by proviral stimulation, a lower abundance of CTLs (see Figure 6.10), a notion which is consistent with current understanding of host cellular immunity.

The composition of the virus, however, was found to be differentially influenced by CTL efficiency in each of the two regions of strain structure (see Figures 6.11(a)–(b)). Specifically, within the DSS region, the fittest strain was always observed to be dominant[4] (at least in the case we have considered here in which only strain 11 exhibits a fitness advantage over the other strains), and such dominance was highly robust and resilient to changes for a broad range of skewness in CTL strengths (Figure 6.12(a)). In contrast, the composition of the virus was not well-preserved in the PDSS region and could vary substantially depending on the differences between the epitope-specific CTL responses. Within this region of PDSS, the fittest strain could range from being the least abundant to the most abundant (Figure 6.12(b)).

Significantly, these results demonstrate how functional differences between the two types of CTL responses can differentially influence the composition of the virus population, especially in the PDSS region, a feature not apparent simply by looking at the magnitude of the proviral load. To illustrate this, we refer to Figure 6.11(b), which shows the abundances of each strain for varying combinations of Tax- and HBZ-specific CTL efficiencies (selected

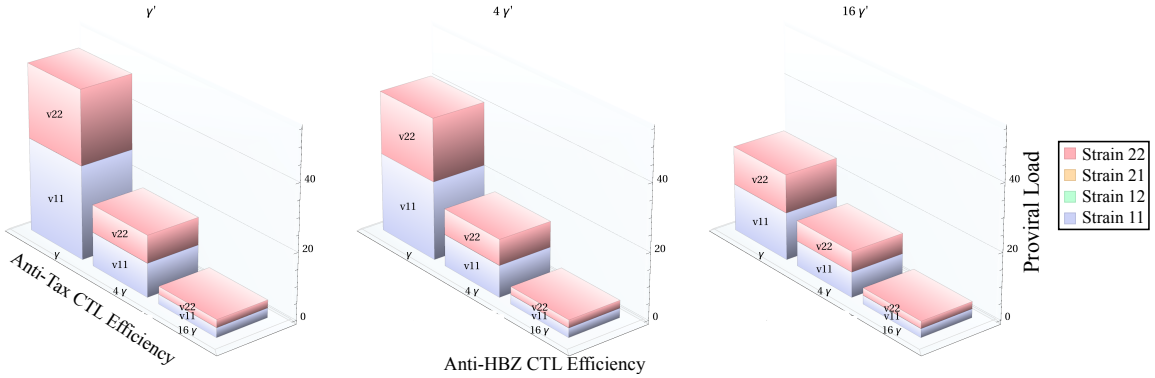
[4]We consider a strain to be dominant if it is the most abundant amongst all of the strains present.



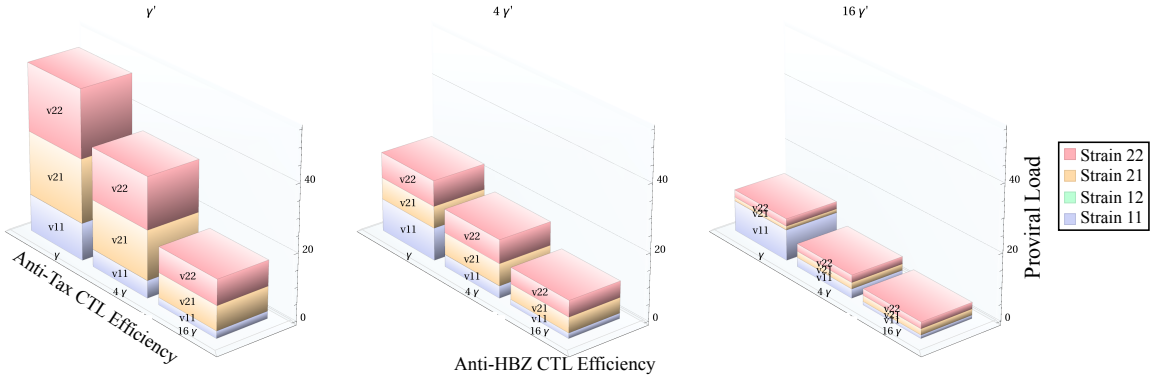
(a) Proviral Load. (b) CTL Frequency.

Figure 6.10: The impact of CTL efficiency on proviral load and CTL frequency for a typical choice of δ_1 and δ_2 in the PDSS region. The two figures show the outcome of solving the multi-locus model (6.1) numerically for each combination of selected rates of CTL lysis, or efficiencies, of both Tax-specific and HBZ-specific CTLs, and plotting the corresponding magnitudes of (a) the proviral load, and (b) the total CTL frequency. Here, we are considering the 2×2 model, where $R_{0,11} = 2.71$ and $R_{0,\text{others}} = 2.56$. Meanwhile, we select multiples of 1, 4 and 16 of the baseline values for the efficiencies, γ and γ', respectively, of Tax-specific and HBZ-specific CTL types. The above graphs correspond to particular points in the plots shown in Figure 6.9. The specific choices for the immunogenicities of the two Tax variants are $\delta_1 = 0.9$ and $\delta_2 = 0.1$, which lies in the PDSS region for all combinations of γ and γ' (refer to Figure 6.9). All other parameter values have been selected from biologically realistic ranges as discussed in Table 3.1. The above figures are representative of any other choice of δ_1 and δ_2 in the PDSS region. However, further investigations have revealed that both the proviral load and frequency of CTLs are negatively correlated with CTL efficiency in the DSS region as well (results not shown).

from multiples of the baseline values of γ and γ' as listed in Table 3.1), with all other parameters fixed, for a typical plot in the PDSS region of the 2×2 model. Specifically, we look at the two bar charts on opposite corners indicated by a choice of $(16\gamma, \gamma')$ (Figure 6.11(b), left-most panel, third bar chart) and the choice of $(\gamma, 16\gamma')$ (Figure 6.11(b), right-most panel, first bar chart) for the respective efficiencies of Tax- and HBZ-specific CTLs. We observe that both cases result in similar total virus abundances (i.e. the heights of the respective bar charts are similar), yet they exhibit vastly different virus compositions: Whereas the former case consists of a relatively uniform mix of strains 11, 21, and 22, the latter case exhibits dominance primarily of strain 11. Since the only differences between the various bar charts are in the efficiencies of the CTL responses, the above observation may have relevance to the diagnosis of HTLV-I-associated disease as well as vaccine administration, which we will discuss in more detail later in this chapter.



(a) Proviral load and strain structure in the DSS region.

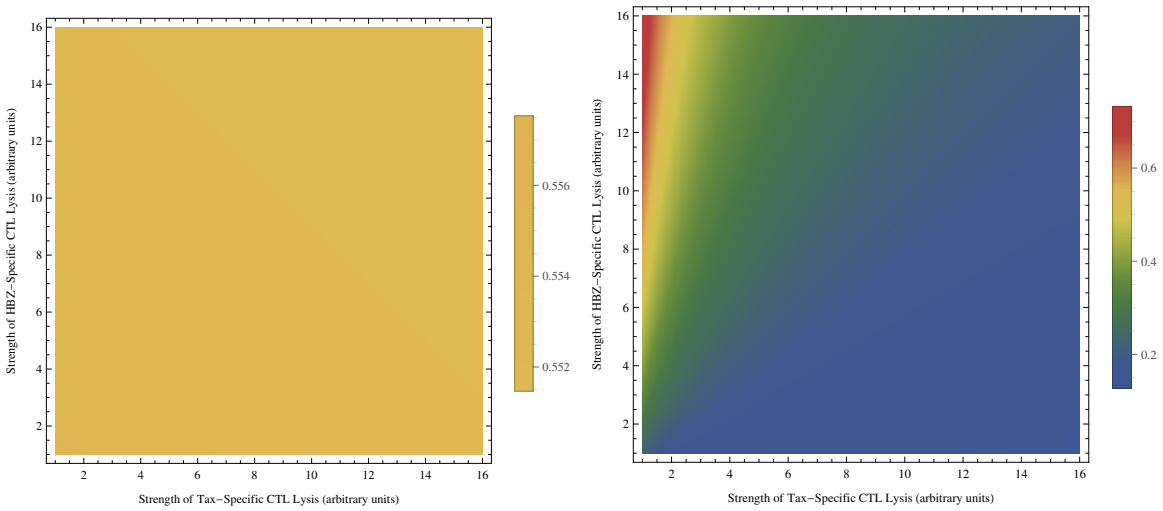


(b) Proviral load and strain structure in the PDSS region.

Figure 6.11: The influence of CTL efficiency on the magnitude of the proviral load and viral strain composition in the DSS and PDSS regions. Our results show that efficient immune responses of either specificity reduce the magnitude of the proviral load regardless of the type of strain pattern exhibited. The composition of the virus population, however, varies between the two regions: (a) Strain composition is well-preserved in the DSS region and is robust to variations in the strength among virus-specific CTLs. (b) By contrast, in the PDSS region, differences in efficiency between the two epitope-specific CTLs can affect the relative abundances of strains within the viral structure. The above figures are representative of the behaviour observed in the DSS and PDSS regions for the 2×2 model. Here, the fitnesses for each strain are as follows: $R_{0,11} = 2.94$ and $R_{0,\text{others}} = 2.56$. The specific choices for the Tax variant immunogenicities for the above figures are as follows: DSS region, $\delta_1 = 0.9$, $\delta_2 = 0.9$; PDSS region, $\delta_1 = 0.9$, $\delta_2 = 0.1$. All parameters have been taken from biologically relevant ranges as in Table 3.1.

6.5 Visualisation of Strain Structure in a Higher Order Multi-Locus System: The 3×3 Model

Because of the great deal of complexity, both analytical and numerical, inherent in the general multi-locus model (6.1), so far we have concentrated on highlighting the basic principles of the multi-locus HTLV-I system by considering only the 2×2 case. Indeed, there arise several challenges when attempting to explore the model with a larger number of possible



(a) Dominance of most advantageous strain is unaffected in the DSS region.

(b) Dominance of most advantageous strain can vary in the PDSS region.

Figure 6.12: Heat map showing the effect of CTL efficiency on dominance of the fittest strain. The above figures illustrate the relative prevalence of the most advantageous strain scaled with respect to the total proviral load and hence represent the level of dominance of the fittest strain within the viral population. (a) In the DSS region, the fittest strain not only preserves dominance, but also maintains a largely invariable relative abundance in the presence of differences in efficiencies between the two CTL types. (b) Meanwhile, in the PDSS region, the dominance of the fittest strain is more variable. Indeed, depending on the efficiencies of each CTL response, the most advantageous strain can either be sub-dominant (i.e. not dominant), weakly dominant (i.e. is the most prevalent strain, but other strains co-exist at significant abundances), or strongly dominant (i.e. makes up almost all of the strain population at equilibrium). The above heat maps are representative of the behaviour observed in the DSS and PDSS regions for the 2×2 model. Strain fitnesses: $R_{0,11} = 2.94$ and $R_{0,\text{others}} = 2.56$. Tax variant immunogenicities: DSS region, $\delta_1 = 0.9$, $\delta_2 = 0.9$; PDSS region, $\delta_1 = 0.9$, $\delta_2 = 0.1$. All parameters have been taken from biologically relevant ranges as in Table 3.1.

variants at each epitope. These include, for instance, a rapid rise in dimensionality of the non-linear system of equations in model (6.1), as well as how we can represent the model results in a clear and concise way that can highlight the relationship between strain composition and immunogenicity of Tax variants which, from the 2×2 case, has been shown to play a substantial role in establishing stable strain structure.

In Figure 6.13, we illustrate a schematic that we can use to visualise the composition of the virus population for a specified combination of Tax variant immunogenicities in the case where there are three variants present at each of the two antigenic loci; in other words, the 3×3 model, which is obtained when $m = n = 3$ in model (6.1). Figures 6.14, 6.15 and 6.16 demonstrate three sample outcomes from numerical simulation of the 3×3 model, which

illustrate the emergence of organised strain structure, both DSS and PDSS. These preliminary results, although representing the typical sort of behaviour exhibited by the model, simultaneously emphasise the need for further investigation in order to better characterise the general multi-locus model with an arbitrary number of variants in both epitopes.

Remark 6.6. The multi-locus modelling approach that we have developed in this chapter can also be applied to the situation in which there is an asymmetric number of variants in each of the two epitopes. As with the case where both epitopes have a symmetric number of protein variants, there exist a wide range of possible overlapping and non-overlapping strain structures. As an example, we refer the reader to Appendix C.4 for representative graphs for the scenario where there are 3 variants in Tax, but only 2 variants in HBZ.

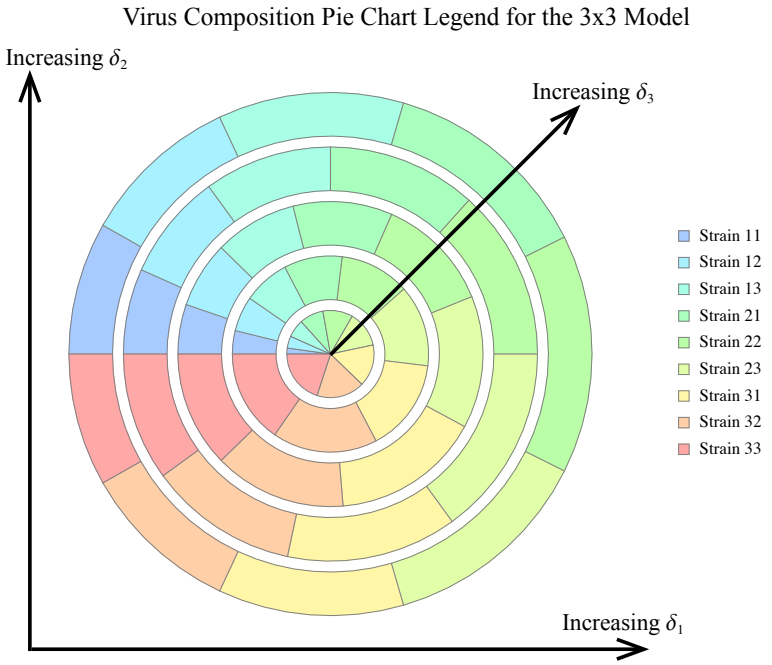


Figure 6.13: Pie chart legend indicating virus composition in terms of relative abundances of each individual viral strain for the higher order multi-locus model with three variants in each of Tax and HBZ. Specifically, this schematic legend illustrates how we can use a 'concentric pie chart representation' to visualise the structure of the virus population at equilibrium according to model (6.1) in the case where there are three possible variants at each of the two epitopes, i.e. the 3×3 model. Here, we consider a total of nine possible virus strains which, listed in lexicographic order, are: 11, 12, 13, 21, 22, 23, 31, 32 and 33. In this representation, we consider a series of circular plots (like the one shown above) consisting of a set of concentric pie charts. Each individual pie chart corresponds to a particular combination of selected values for the immunogenicities of each of the three distinct Tax variants; that is, δ_1, δ_2 and δ_3. In the subsequent simulations, we consider five different sample choices, representing a set of evenly spaced values between 0 and 1, for each of the δ_i. Separate circular plots along the horizontal and vertical axes correspond to the various choices of δ_1 and δ_2, respectively (note that the schematic legend above only shows one such circular plot). Meanwhile, within each circular plot, the pie charts spreading radially outward correspond to increasing values of δ_3. For each pie chart, the colours represent individual viral strains as indicated. As we will see in the remainder of this chapter, this concentric pie chart representation allows us to examine and gain insights to the way in which Tax variant immunogenicities influence the composition of the virus when the various strains all differ in their respective levels of fitness (in terms of the basic reproduction number $R_{0,ij}$ for strain ij). Note the slight difference in the interpretation of the concentric pie charts radiating outwards in this representation versus those described earlier in Figure 6.6.

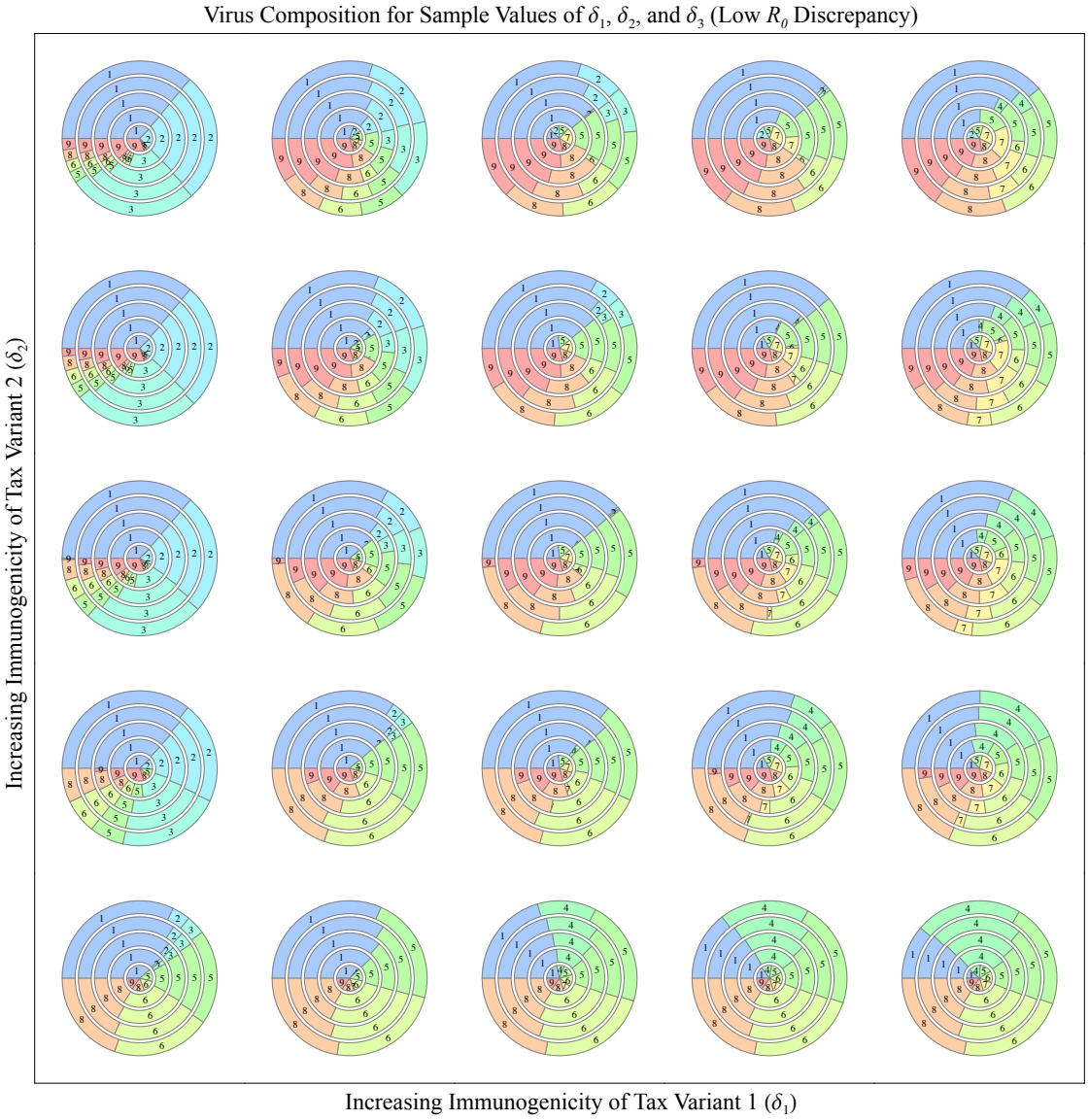


Figure 6.14: Visualisation of virus composition at equilibrium in the 3×3 multi-locus model. The above figure shows one possible outcome of solving model (6.1) with $m = n = 3$ numerically, and plotting the relative abundances of all co-existing virus strains at equilibrium for a particular combination of Tax variant immunogenicities (i.e. δ_1, δ_2, and δ_3). Strains are labelled 1 through 9 corresponding to the strains ij in lexicographic order. In this figure, only strain 11 (labelled '1') is different from all of the others. There is no DSS present in this scenario, which is expected since all other strains aside from strain 11 (labelled '1') have the same fitness. Nevertheless, the strain structure consisting of strains {11, 22, 23, 32, 33} (labelled {'1', '5', '6', '8', '9'}, respectively) is similiar to DSS. Here, $R_{0,11} = 2.94$, $R_{0,\text{others}} = 2.56$. Each of the Tax variants can take on one of the following five (uniformly-spaced) immunogenicity values between 0 and 1: 0.05, 0.275, 0.5, 0.725, 0.95. Thus, a total of $5 \times 5 \times 5 = 125$ simulations are represented. All other parameters have been selected from Table 3.1.

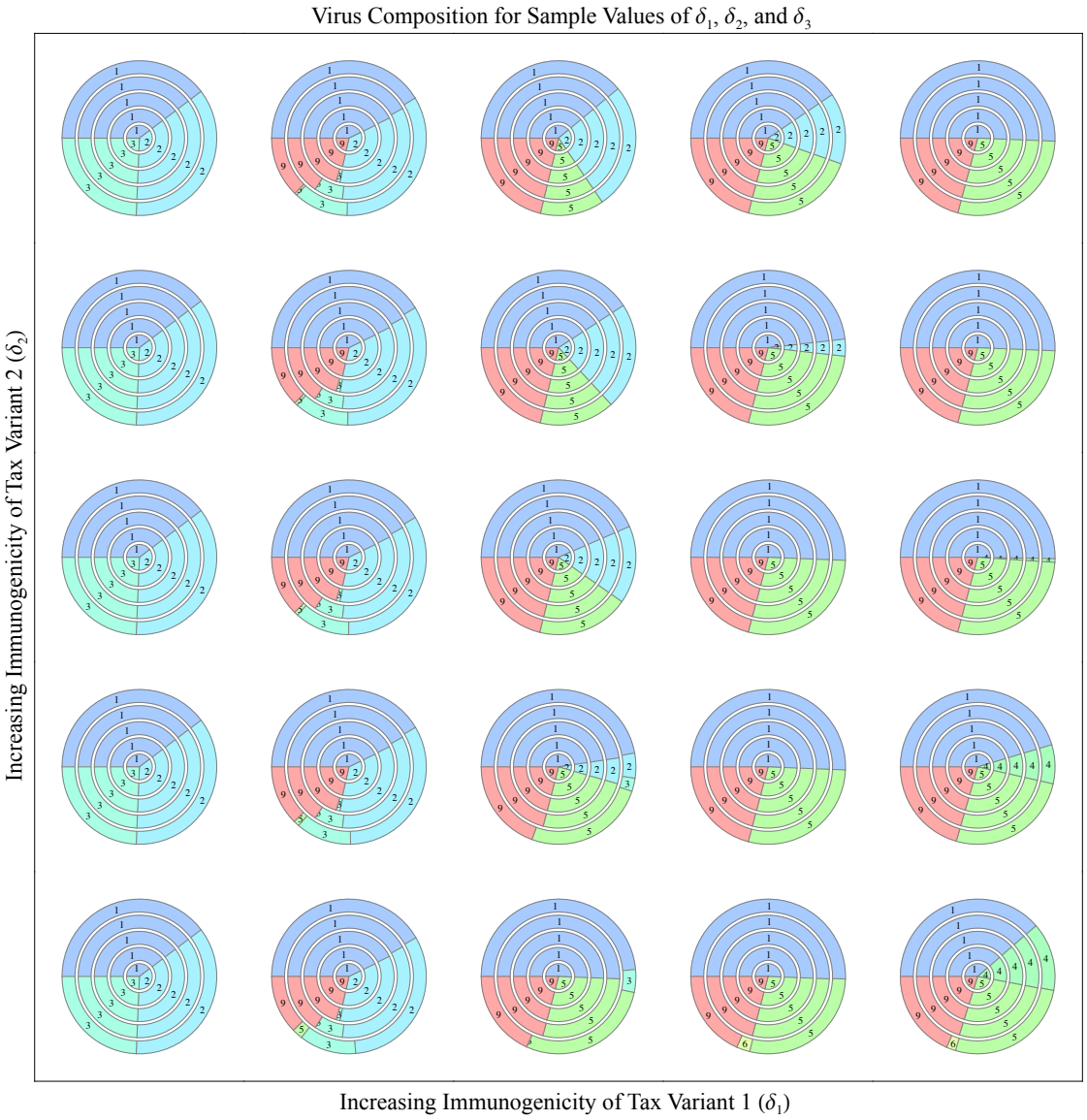


Figure 6.15: DSS and a variety of overlapping strain structures can exist in the 3×3 system. The above figure shows the result of numerical simulation of the multi-locus model (6.1) with $m = n = 3$. In particular, we demonstrate that it is possible for DSS to appear in the 3×3 model, at least for the given set of parameters considered here and particular combinations of Tax variant immunogenicities. In this case, when DSS emerges, the virus population consists of the strains {11, 22, 33} (labelled {'1', '5', '9'}) and, moreover, the fittest strain, which is strain 11, is always dominant in that it exists at the highest prevalence among all existing strains. Here, the fitnesses of all strains have been randomly assigned and the basic reproduction numbers for each strain are $R_{0,11} = 10.48$, $R_{0,12} = 9.38$, $R_{0,13} = 6.71$, $R_{0,21} = 4.16$, $R_{0,22} = 5.63$, $R_{0,23} = 2.32$, $R_{0,31} = 1.76$, $R_{0,32} = 5.20$, $R_{0,33} = 4.43$. Each of δ_1, δ_2, and δ_3 can take one of the following five (uniformly-spaced) immunogenicity values between 0 and 1: 0.05, 0.275, 0.5, 0.725, 0.95. Thus, the figure represents a total of $5 \times 5 \times 5 = 125$ model simulations. All parameter values have been selected from physiologically realistic ranges as discussed in Table 3.1.

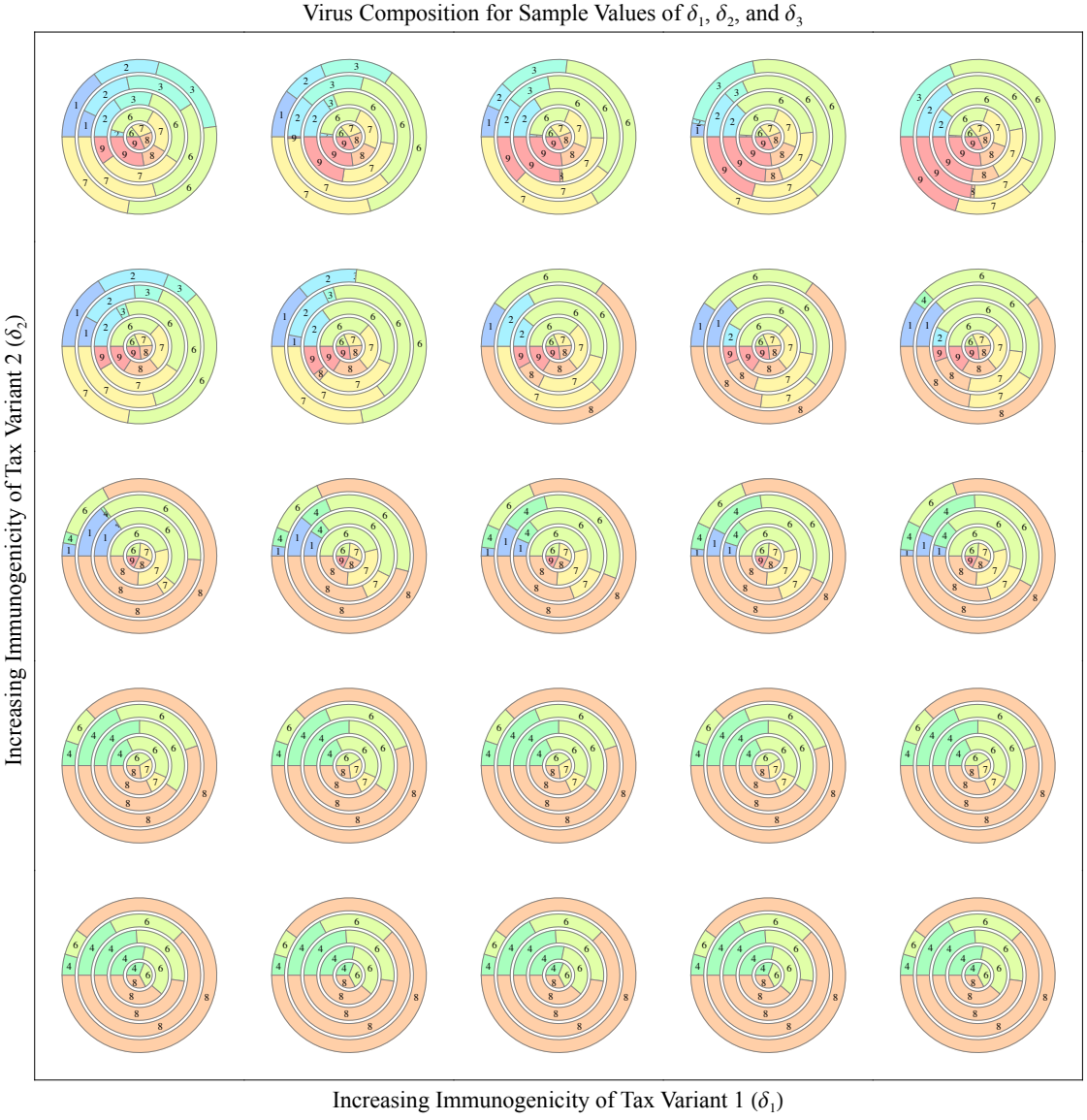


Figure 6.16: Unlike in the 2×2 model, DSS does not imply that the fittest strain is dominant in the 3×3 system. The above figure illustrates yet another possible outcome from numerical exploration of the multi-locus model (6.1) with $m = n = 3$. In contrast to the sample simulation in Figure 6.15, here we observe that for many values of Tax variant immunogenicity δ_1, δ_2 and δ_3, the dominant strain is Strain 32 (labelled '8'), which is not the fittest strain but, in fact, the second fittest strain. Furthermore, there exists a DSS with the strains {11, 23, 32} (labelled {'1', '6', '8'}). Even in these regions of DSS, the fittest strain 23 (labelled '6') is far less abundant than the less fit strain 32 (labelled '8'). Note that there is also a combination of immunogenicities that gives rise to a different set of DSS with the strains {12, 23, 31} (labelled {'2', '6', '7'}), where the fittest strain is the most prevalent. Here, the fitnesses of all strains have again been randomly assigned and are: $R_{0,11} = 4.63$, $R_{0,12} = 4.23$, $R_{0,13} = 5.39$, $R_{0,21} = 9.13$, $R_{0,22} = 3.88$, $R_{0,23} = 11.05$, $R_{0,31} = 8.23$, $R_{0,32} = 9.67$, $R_{0,33} = 8.87$. Each of the Tax variant immunogenicities can take one of the following five (uniformly-spaced) immunogenicity values: 0.05, 0.275, 0.5, 0.725, 0.95. These are meant to represent the range of values between 0 and 1, and leads to a total of $5 \times 5 \times 5 = 125$ combinations of δ_1, δ_2, and δ_3 (corresponding to each of the 125 individual pie charts illustrated above). All parameter values have been selected from Table 3.1.

6.6 Conclusions and Discussion of the Results from the Multi-Locus Model

In this section, we focus on the application of our results from the multi-locus model of HTLV-I infection and discuss their significance for, and implications on, our understanding of the immunological processes behind viral persistence and pathogenesis. Specifically, we link up our model conclusions with clinical and experimental observations to provide biological justification and qualitative validation of the model.

6.6.1 The Emergence of Ordered Strain Structure in Persistent HTLV-I Infection

Our multi-locus modelling approach demonstrates that inter-strain differences in fitness can give rise to two qualitatively distinct types of organised strain structure: DSS and PDSS. DSS is characterised by reduced strain diversity and emergence of a stable set of strains that do not share common antigenic variants at any CTL epitopes, whereas PDSS is distinguished by a greater degree of strain diversity and convergence to a set of strains that express overlapping variants. In the 2×2 model, we found that a characteristic feature of DSS is that it is dominated by the fittest strain and such dominance is generally resilient to differences in the efficiencies of epitope-specific CTLs. Meanwhile, in the case of PDSS, the fittest strain does not necessarily exhibit dominance over the virus population and strain composition is strongly governed by the combinations of Tax variant immunogenicities as well as the relative skew in lytic potential between the various epitope-specific CTL response. We note, however, that in higher dimensional representations of the multi-locus model (6.1), for example, in the symmetric 3×3 model or the asymmetric 3×2 model, DSS does not necessarily imply that the fittest strain dominates the virus population. It is unclear why this would be the case and warrants further exploration of the model to identify the key distinguishing aspects of each type of organised strain structure.

It is unsurprising that the introduction of heterogeneity in the fitness levels among the virus strains induces a change in the virus population structure. Moreover, what one might expect to see is the emergence of a single dominant strain that has the greatest fitness advantage over all other strains, a pattern which is likely to emerge if all virus strains share a conserved epitope that is highly immunogenic; in other words, the strain that has the highest fitness outcompetes and rapidly eliminates all other strains. However, it is known that HTLV-I, like other RNA retroviruses, does not exist as a single, genetically uniform strain; rather, it is comprised of a set of strains that are closely related yet distinct (Kubota et al., 2007; Niewiesk et al., 1994). The distribution of viral strains is also known to differ significantly both between and within seropositive individuals. This composite nature

of HTLV-I is consistent with the predictions of a multi-locus mathematical model that illustrates an alternative dynamical possibility aside from single strain domination and complete viral heterogeneity whereby all possible strains co-exist, namely, the emergence of organised strain structure comprising a subset of persistent strains.

6.6.2 Why Strain Structure?

A natural question that arises is the following: Why does the pathogen tend to organise itself into heterogeneous groups or clusters rather than forming a single homogeneous pool? Mathematically speaking, it has been demonstrated that the formation of ordered sets of strains is an intrinsic feature of mathematical models which describe multi-strain infections displaying a measure of cross-immunity (Calvez et al., 2005; Gupta et al., 1996, 1998; Recker et al., 2004). In our multi-locus model (6.1), the consideration of virus strains that share common antigenic variants at either one of the two distinct epitopes introduces a kind of intrinsic cross-immunity. Indeed, a CTL response that is specific to a particular protein variant (at either epitope) is able to recognise, and therefore suppress, any virus strain expressing the same protein at the same epitope.

From a biological point of view, there are two lines of evidence, which are not mutually exclusive, arising from examination of clinical studies on chronic infection that may help to explain this phenomenon. First, the existence of a vigorous and highly specific anti-HTLV-I CTL response in the chronic phase of infection places continual positive selection pressure for the maintenance of viral antigenic variants. The most direct form of evidence for this comes from a study by Niewiesk et al. (1995), who demonstrated that naturally occurring variants in the Tax protein are able to escape recognition by CTLs specific to the consensus Tax sequence, thereby supporting a role for the establishment and maintenance of viral diversification in concert with the immune selection process. Second, it is believed that the long-term persistence and evolutionary survival of many pathogenic agents, including viruses, are not dependent on the development of pathological conditions (Gupta and Maiden, 2001), and HTLV-I appears to be no exception. The tendency of HTLV-I to self-organise into ordered strain structure promotes the co-existence of distinct viral strains, thereby maintaining pathogen diversity and potentially conferring protection from disease, and could therefore help explain the observation that over 90% of HTLV-I-infected individuals remain as asymptomatic carriers of the virus, while only a small proportion of seropositive individuals develop some form of HTLV-I-associated pathology (Asquith and Bangham, 2007; Proietti et al., 2005).

6.6.3 Virus Composition and Diagnosis of HTLV-I-Associated Pathology

Even though there is a large degree of diversity in viral strains, no single HTLV-I strain is known to be directly associated with disease outcome (Daenke et al., 1990; Mosley et al., 2005; Niewiesk and Bangham, 1996). Nevertheless, despite the intrinsically heterogeneous nature of the virus, the strain that dominates the virus population is likely to exert the greatest influence over the course of infection. Often it is assumed that the strain possessing the highest fitness is the one that has the greatest potential to cause pathological damage to the host. It is therefore plausible that the successful establishment of strong dominance by such a strain in an environment with low strain diversity would be detrimental to the host and may, in turn, be linked to the development of disease. Conversely, suppression of the fittest strain in an environment of higher strain diversity may be protective and indicative of asymptomatic clinical status. In the context of our multi-locus mathematical model, we have identified two distinct types of viral strain structure, DSS and PDSS, each of which is characterised by a set of distinguishable properties. While DSS is stable to variation in CTL efficiencies, PDSS is more susceptible to such factors. We hypothesise that single strain dominance resulting in increased virus homogeneity corresponds to disease manifestation (for example, dominance of the fittest strain exhibited by DSS, or strong dominance by a single lesser-fit strain coupled with a large skew in the efficiencies of host immune responses in certain regions of PDSS), whereas the greater degree of strain diversity along with non-dominance of viral strains generally, but not always, associated with PDSS corresponds to asymptomatic carriage. Therefore, if it is indeed the structure of the virus population, rather than the size, that determines the outcome of disease, then our results suggest a way to diagnose the clinical status of infected individuals irrespective of the proviral load. Specifically, heterogeneous strain structure is associated with asymptomatic clinical status and protection from disease, while homogeneous strain structure is associated with HTLV-I-associated pathology, such as HAM/TSP and ATL.

The notion that the structure of the pathogen population, and not necessarily the presence of any specific strain nor the size of the proviral load, is linked to disease progression has been suggested by clinical evidence. Indeed, an experimental study by Niewiesk et al. (1994) on the intra-isolate variability of the transcriptional transactivator Tax provides support for such a hypothesis, demonstrating that healthy HTLV-I seropositive individuals harbour virus populations displaying wider variation at the protein level than do patients with HAM/TSP and indicating that a greater degree of amino acid conservation is associated with more severe pathological manifestation.

Although the precise threshold of viral homogeneity or heterogeneity associated with disease manifestation is currently unknown and, moreover, is largely dependent on both

virus and host factors, it is hoped that with the increase in accuracy and availability of longitudinal data, it may be possible to quantify such thresholds and verify the plausibility of this hypothesis. In terms of testing and validation, relevant data could be obtained from experimental studies that quantify the sequence variation in both Tax and HBZ for a given sample as well as identify the clinical status of the HTLV-I-infected subject. As discussed above, this has been carried out for the Tax protein (Niewiesk et al., 1994); however, the potential for simultaneous expression of distinct HBZ variants, even with the same Tax variant, would imply a range of different 'strains', at least in the context of our multi-locus modelling framework. Such a study could help test for the association between different 'strains' and clinical outcome and, if it exists, refine our hypothesis regarding strain heterogeneity and risk of developing disease.

Remark 6.7. The concept of heterogeneity in relation to disease progression has been explored by experimental groups in differing, but analogous, contexts. For instance, Gillet et al. (2011) have recently developed an experimental technique to characterise the abundance of infected T-cell 'clones' within the HTLV-I-infected T-cell population, where each clone has been identified by the unique insertion site of the provirus in the host genome. The authors devise a metric based on the Gini coefficient[5], called the oligoclonality index, to quantify the clonality of a given sample, that is, the distribution of the abundance of each clone. Using this metric, the authors make the remarkable observation that the proviral population in infected patients with malignant disease (the various manifestations of ATL) is comprised principally of a relatively small handful of clone types that exist in large abundances and account for major proportions of the virus population. These findings suggested that a more homogeneous distribution of HTLV-I clones was correlated with severe pathology — for instance, Gillet et al. (2011) reported that in a patient with chronic leukaemia, more than 80% of the HTLV-I-infected target cell population was made up of a single, abundant clone. By contrast, the distribution of infected clones in HTLV-I seropositive individuals without malignant disease showed a substantially greater degree of uniformity among the viral strains, although the authors remarked that the oligoclonality index could not distinguish ACs from HAM/TSP patients. Lastly, we note that despite the difference in context from our multi-locus model, where we consider the relationship between disease progression and heterogeneity in a set of distinct virus 'strains' and not 'clones', the study by Gillet et al. (2011) nevertheless provides further support to the notion that a reduction in diversity within the pathogen structure may be associated with HTLV-I-linked disease,

[5]The Gini coefficient (Gini, 1912) was originally defined to be a measure of inequality in the distribution of wealth or income of individuals.

although it is important to note that in both situations such an association demonstrates a correlation and is not indicative of a causal relationship.

6.6.4 Both Virus and Host Factors Contribute to the Establishment and Maintenance of Strain Structure

HTLV-I-associated pathology often manifests in seropositive individuals after a long asymptomatic phase. Although the incidence of severe HTLV-I-associated pathology is low, the fact remains that disease still manifests in a handful of individuals. What factors could cause pathogenesis? Based on our hypothesis that viral strain structure is a determinant of disease, our model results demonstrate the way in which each side of the host-virus interaction could aid in establishing organised strain patterning.

With respect to viral factors, our model suggests that a large fitness discrepancy between the most advantageous strain and the remainder of the virus population promotes the establishment of a discrete strain structure (DSS) as well as strong dominance of the fittest strain, reducing viral heterogeneity and thus implying an increased risk of disease acquisition. These inter-strain differences may arise from fixation of certain strains which in some way have a substantial fitness advantage over others, for instance in the activity of transcriptional transactivation. Such a scenario could potentially explain the association between four nucleotide substitutions in the *tax* gene and a higher incidence of HAM/TSP in a population of HTLV-I seropositive individuals with and without the inflammatory disease HAM/TSP (Furukawa et al., 2000; Nagai and Osame, 2003). For instance, the nucleotide substitutions may have created a larger fitness discrepancy among the various strains, leading to an increased likelihood of pathology, although this is speculative.

Despite exhibiting a significant degree of inter-strain variability, the HTLV-I genome is relatively conserved compared to other retroviruses and differs little in sequence between infected individuals (Mortreux et al., 2003; Wattel et al., 1996). Such conservation is due primarily to functional constraints on viral proteins. For instance, many (but not all) amino acid substitutions in the Tax protein abolish its transactivation ability (Niewiesk et al., 1995). Nevertheless, HTLV-I-infected individuals present a broad range of clinically relevant data, including the proviral load, the structure and relative abundance of individual strains within the viral population, and the existence and severity of associated pathologies. Due to the genetic stability of HTLV-I, inter-strain variation alone is unlikely to be able to explain the between-host variability that is observed and it is therefore believed that differences among virus-infected hosts, and not only in the virus itself, can play an integral role in determining clinical status.

At the genomic level, hosts differ in their HLA type, which is a genetic factor governing the functionality of their immune responses against invasive pathogens, including viruses. There is mounting evidence that such host genetic factors strongly influence the course and outcome of chronic HTLV-I infection, and that certain HLA class I alleles, which govern the specificity and functionality of cellular immune responses, can be classified as being protective or detrimental. Indeed, in a cohort of HTLV-I seropositive individuals in southern Japan, a lower proviral load as well as a reduced prevalence of HAM/TSP was observed in those carrying either one of the HLA class I alleles *HLA-A*0201* and *HLA-Cw*0801*; meanwhile, the HLA class I allele *HLA-B*5401* was associated with an increased proviral load and greater risk of developing inflammatory disease (Bangham, 2003; Jeffery et al., 1999, 2000; Vine et al., 2002). In a geographically similar cohort, *HLA-A*26*, *HLA-B*4002*, *HLA-B*4006*, and *HLA-B*4801* alleles have been linked to the progression of ATL (Yashiki et al., 2001).

An individual's HLA type determines the lytic potential of the HTLV-I-specific CTL response, which is governed by two primary factors: CTL strength or efficiency, which is generally characterised by the rate of cell-mediated lysis of provirus-positive target cells, and effective CTL expansion, which is determined by the proliferation rate and lifespan of effector cells. With respect to our multi-locus model, the role of CTL efficiency is captured in the parameters γ and γ' representing the respective rates of Tax- and HBZ-specific CTL-mediated lysis. Meanwhile, effective CTL expansion of Tax- and HBZ-specific CTLs are given by the parameters ϕ_{Tax} and ϕ_{HBZ}. In addition, the activation of virus-specific CTLs by HTLV-I antigens not only triggers their rapid expansion, but also initiates activation-induced programmed cell death, a process which shortens the lifespan of such effector cells and therefore limits their cumulative potential for eliminating proviral cells. Recently, interest has been focussed on the role of killer cell immunoglobulin-like receptors (KIRs) in influencing the course and outcome of chronic viral infections such as HTLV-I (Seich al Basatena et al., 2011; Ugolini et al., 2001). In particular, experimental evidence has suggested that a subset of KIRs, known as inhibitory KIRs, usually associated with innate immunity, can regulate the adaptive immune response, primarily by extending CTL survival and circumventing activation-induced cell death. It has been proposed that KIRs extend CTL survival in an environment of chronic antigen stimulation by raising the threshold for CTL activation (Ugolini et al., 2001).

We discovered that a skew between the lytic potentials of different epitope-specific CTL responses changes the way in which Tax variant immunogenicities give rise to each type of organised strain structure. Specifically, stable DSS was favoured for greater immunogenicity of the Tax variant carried by the fittest strain when CTL lytic potential was skewed towards

the HBZ epitope (i.e. $\gamma/\gamma' < 1$ or $\phi_{\text{Tax}}/\phi_{\text{HBZ}} < 1$). Conversely, when CTL lytic potential was skewed towards the Tax epitope ($\gamma/\gamma' > 1$ or $\phi_{\text{Tax}}/\phi_{\text{HBZ}} > 1$), lower immunogenicity of the Tax variant carried by the fittest strain was seen to promote DSS. Investigation of virus and CTL structure with respect to differences in the lifespans of CTLs showed similar shifts in how immunogenicity induces each type of strain pattern.

We suggest that protective HLA class I alleles change the host immune landscape in favour of heterogeneous viral strain structure (i.e. PDSS), resulting in suppression of strains that have too much of a fitness advantage over the others, whereas HLA class I alleles associated with detrimental effects shape the immune environment towards viral strain structure that is homogeneous and dominated by a particularly fit strain (i.e. DSS). Experimental studies to test such a hypothesis would require classification of HTLV-I sero-positive individuals according to their HLA class I alleles, clinical status (e.g. asymptomatic, HAM/TSP, ATL), as well as measurements of the proviral population collected from blood sera to identify both strain structure and the fitnesses of each individual strain. This would allow us, firstly, to assess the hierarchy of the proviral population and verify whether or not an association exists between the type of strain structure possessed by an infected individual and that particular individual's HLA class I allele and, secondly, to determine the link with the presence (or absence) of disease manifestation.

6.6.5 Implications for Administration of Vaccines

Lastly, our work has significant implications regarding the administration of anti-viral vaccine treatments. Vaccines can commonly be designed to improve the efficiency of host immunity with the aim of reducing total virus abundance. Typically, such vaccines are highly specific to target antigens (in other words, they are not broadly-neutralising). With respect to our multi-locus model (6.1), we discovered that although boosting the efficiency of either CTL response, Tax-specific or HBZ-specific, can effect a similiar reduction in the size of the proviral load (see Figure 6.10(a)), our results show that administering one type of vaccine focussed on one of the epitopes could reduce the proviral load without substantially changing viral strain structure, whereas administration of another vaccine targeting the other epitope could achieve a similar reduction in the proviral load, but significantly disrupt the structure of the virus, making it more homogeneous and therefore increasing the infected individual's risk of developing disease (refer to Figure 6.11). Because of the potential impact on vaccine design and administration of treatment regimes, further experimental and theoretical work is required in order to assess the feasibility of this hypothesis.

6.7 Final Remarks on the Multi-Locus Model

The consideration of simultaneous expression of multiple antigenically variable HTLV-I antigens, namely Tax and HBZ, accompanied by variant-specific CTL responses adds a layer of complexity to the way we think about the host-virus dynamics in HTLV-I infection. We examined these interactions in the context of a multi-locus mathematical modelling framework. Due to the added mathematical complexity arising from the incorporation of more detailed immunological interactions, a complete analytical characterisation of the general multi-locus model was not possible. We therefore focussed our attention on the 2×2 case and explored the dynamical behaviour of this model extensively using computational methods in order to gain a sense of the basic underlying features of the multi-locus HTLV-I system. Preliminary results for a higher dimensional system exhibiting symmetry (i.e. the 3×3 model) and asymmetry (i.e. the 3×2 model) in the number of variants at each antigenic locus, Tax or HBZ, appear to demonstrate similar types of qualitative behaviour with respect to the emergence of ordered strain structure. Nevertheless, further investigation is required to fully understand the details of the general formulation of the multi-locus model of HTLV-I infection.

The work in this chapter may be relevant to research on the within-host dynamics of persistent infection by the human immunodeficiency virus (HIV), an antigenically variable infectious pathogen that is much better-studied and for which a greater amount of data is available. Like HTLV-I, chronic HIV infection is characterised by a virus population made up of a collection of distinct virus variants (Goodenow et al., 1989). However, HIV displays a substantially greater degree of genetic diversity than HTLV-I, and such rapid evolution of HIV results in the emergence of escape mutants which successfully evade host immunity (Goodenow et al., 1989; Rambaut et al., 2004; Smyth et al., 2012). In contrast, HTLV-I exhibits a lower degree of sequence diversity (Bangham, 2000b), meaning that mutation is likely to play a less significant role for HTLV-I than HIV. An understanding of the HTLV-I system can help unravel the key underlying principles governing the host-virus interaction between a persistent human retrovirus and host immunity in the (relative) absence of escape mutations. These basic underlying principles could then form a conceptual basis when carried over to HIV infection, where escape mutations are more prevalent. Considering the wealth of experimental and theoretical work done on HIV, including studies on identifying the association between HLA type (which is involved in CTL restriction) and the outcome of the infection (Asquith et al., 2006; Asquith, 2008; Leslie et al., 2004), as well as mathematical modelling of within-host HIV infection (Fryer et al., 2010; Nowak et al.,

1995a,b; Nowak and Bangham, 1996; Perelson et al., 1993; Perelson, 2002), a further exploration and connection between HTLV-I and HIV research can lead to a better understanding of the role of host immunity against both of these persistent retroviral infections.

Chapter 7

Conclusions and Future Directions

7.1 Perspectives and Overview

Understanding the complex mechanisms that HTLV-I has developed to evade elimination despite the presence of a strong, virus-specific host immune response is an issue that has puzzled theoreticians and clinicians. We explored this issue of HTLV-I persistence from a mathematical modelling point of view. Using a non-linear dynamical systems approach, our goal in this thesis has been to develop a consistent mathematical framework that can capture the essential features of the dynamic host-virus interactions that take place in HTLV-I infection. Specifically, we constructed a series of four mathematical models aimed at progressively incorporating three broad themes: (I) proviral latency and activation, (II) simultaneous viral protein expression, and (III) antigenic variability. Although the models each focus on a particular biological feature of interest, they are formulated so as to maintain aspects of previous themes whilst exploring subsequent themes; that is to say, the ideas developed in Theme I follow through to Themes II/III and, similarly, the ideas of Theme II form a basis for examination of Theme III (refer to Figure 7.1 for a schematic diagram of how each of the mathematical models in this thesis falls within the three themes). By analysing the resulting mathematical models, we have attempted to extract insights that can help elucidate the immunological phenomena of HTLV-I persistence amidst host immunity and the manifestation of HTLV-I-associated diseases. In this chapter, we re-visit and summarise the outcome of our investigations regarding each of the three themes and finish off with several short concluding remarks.

Theme I: Proviral Latency and Activation

The basis of the work presented in this thesis has been motivated by the novel hypothesis by Asquith and Bangham (2008), who proposed that HTLV-I persistence is the consequence of a dynamic balance between proviral latency and activation. We developed this idea in

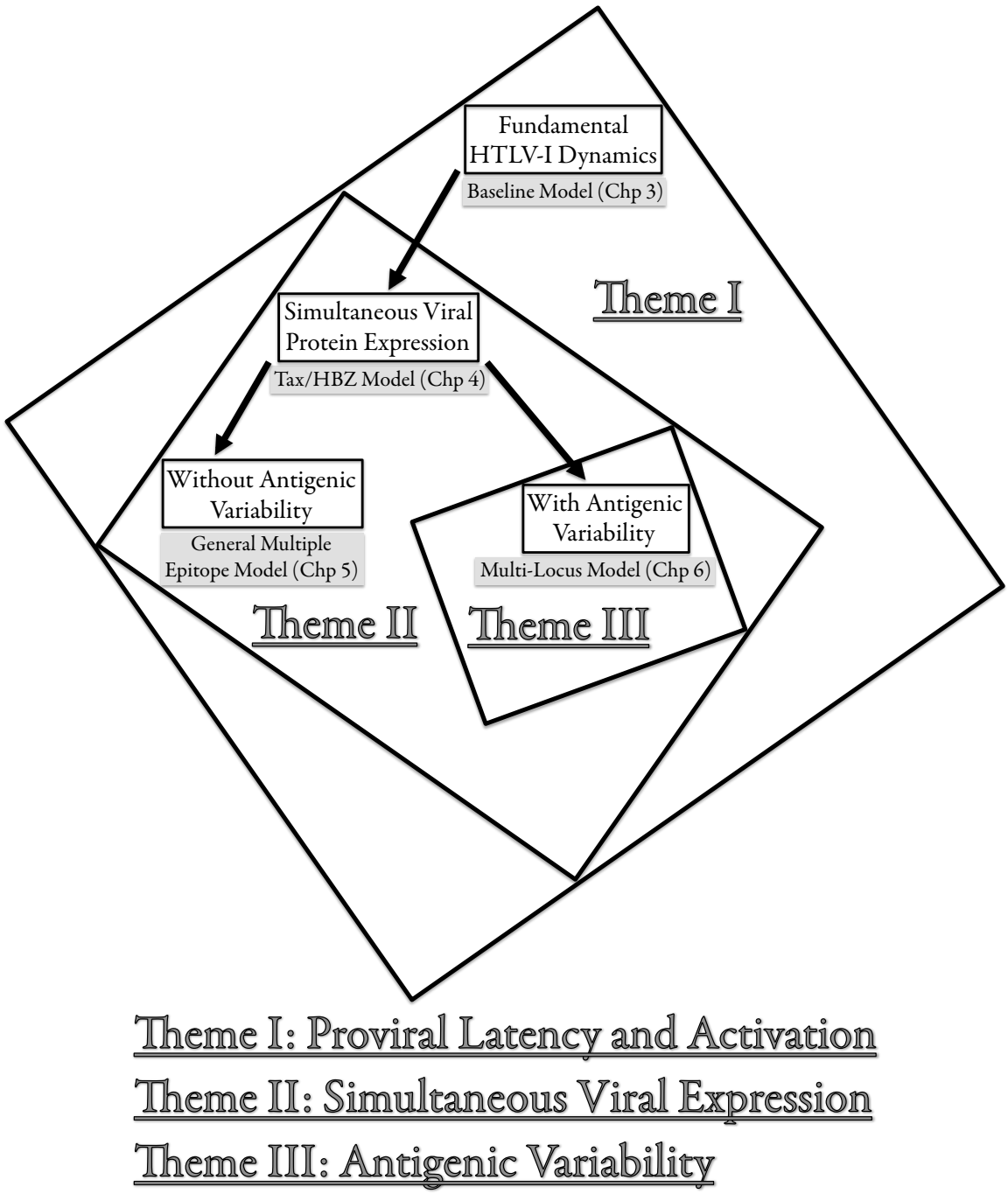


Figure 7.1: Overview of the three broad research themes explored in this thesis and how each of the mathematical models relate to them. We have constructed and analysed a series of four models that progressively develop three broad themes: (I) Proviral latency and activation, (II) simultaneous viral expression, and (III) antigenic variability.

mathematical terms by formulating and analysing a 'baseline model' in Chapter 3, which focussed on exploring the role of proviral latency and activation (Theme I) during chronic HTLV-I infection. The results of the baseline model support the conclusions of Asquith

and Bangham (2008) and help justify why HTLV-I is not a silent infection, what determines immune efficiency, and why infection is life-long. Namely, our model demonstrated that viral expression is necessary for the establishment, transmission, and persistence of HTLV-I, that CTL efficiency depends not solely on the abundance of effector cells, but also substantially on the rate of cell-mediated lysis of active proviral cells, and that the reason infection is difficult to clear is because the population of latent proviral cells supply the pool of activated proviral cells that aggressively propagate the infection (due to continuous spontaneous proviral activation). This model is consistent with the relatively recent idea that the extent of proviral activation, rather than the magnitude of the proviral load, may be a strong determinant of clinical status (i.e. the presence or absence of disease) among HTLV-I seropositive individuals, a notion which is supported by experimental evidence (Asquith and Bangham, 2008). Moreover, interpretation of the model results allows us to propose a plausible mechanism for HTLV-I pathogenesis. This baseline model forms a foundation for the remaining models in subsequent chapters.

Theme II: Simultaneous Viral Expression

Having established a solid foundation in the baseline model, we set out to incorporate a greater degree of immunological detail, namely, the simultaneous expression of multiple viral antigens (Theme II), a feature that has not been previously considered in mathematical models of HTLV-I infection. We first considered the case where there are two epitopes that are displayed by a proviral cell, leading to what we refer to as the 'Tax/HBZ model' in Chapter 4, aptly named as such to emphasise that Tax and HBZ are the two key viral proteins that are believed to play the most significant role in the establishment, persistence, and pathogenesis of infection by HTLV-I (Boxus and Willems, 2009; MacNamara et al., 2010; Matsuoka and Green, 2009), although our approach does not explicitly distinguish between the two distinct proteins with respect to functionality. This basic two-epitope model shifted the focus to the way that HTLV-I-infected target cells are recognised by host immunity, which has evolved to be highly specific to target antigens, and emphasised the intrinsic heterogeneity present in the CTL population. In order to identify how each type of CTL, Tax-specific or HBZ-specific, could impact the course of infection, we determined the principal factors that quantify their respective relative abundances within the CTL pool. Our model results demonstrated that the effective rate of expansion (i.e. growth/death ratio) was not the only factor determining the prevalence of each CTL type but, significantly, the relative immunogenicity of its cognate epitope also played a pivotal part in shaping the CTL landscape.

From this point, we branched off and extended the Tax/HBZ model in two separate directions. Along one route, we further expanded the ideas of Theme II by considering the scenario in which an arbitrary number of viral proteins are expressed at the same time, leading to the general 'multiple epitope model' in Chapter 5. Indeed, although Tax and HBZ are believed to be the two principal viral proteins in persistent HTLV-I, it is nevertheless the case that up to 12, or possibly more, HTLV-I proteins are displayed simultaneously by an activated proviral cell (Boxus and Willems, 2009; Goon et al., 2004a; MacNamara et al., 2010), and so we set out to explore this feature in more detail. The results of this model helped to emphasise the significance of relative immunogenicities in shaping the CTL repertoire and demonstrated that effective control of an individual's proviral load is the outcome of an immunogenicity profile which selects for the dominance of CTL types that expand rapidly, are efficient at eliminating proviral cells, and target highly immunogenic epitopes. Furthermore, by incorporating existing published data sets into this general multiple epitope modelling framework, we were able to derive preliminary estimates for the average fitness both of the virus as well as the CTL repertoire within a particular HTLV-I-infected individual. These preliminary estimates are in general agreement with experimental data, however, further investigation is required to establish validity of the model.

Theme III: Antigenic Variability

Along the alternate route, we expanded the Tax/HBZ model in a different direction by taking into consideration the immunological feature of antigenic variability (Theme III) in HTLV-I infection, which arises due to the existence of protein variants in both the Tax and HBZ epitopes (Gaudray et al., 2002; Kubota et al., 2007; Niewiesk et al., 1995). To capture such a feature, we model the HTLV-I system using a multi-locus approach, a theoretical framework which has been successfully applied to a wide variety of infectious pathogens exhibiting some measure of antigenic variability (Gupta et al., 1996, 1998; Buckee et al., 2008; Recker et al., 2007; Nowak et al., 1995a; Wikramaratna et al., 2013).

In HTLV-I, the presence of antigenic variability means that a specific combination of variants in each of Tax and HBZ defines a particular virus subpopulation (which we refer to as a 'strain'). Moreover, the highly-specific nature of host cellular immunity implies that a broad range of CTL responses exist, each of which can only recognise a single variant at a particular epitope (known also as an 'antigenic locus'). We explored this theme via the formulation of a general 'multi-locus model' in Chapter 6, where antigenic variability is considered at the two epitopes. The incorporation of multiple viral strains raises the issue of heterogeneity in the virus population. Using these models, we set out to examine whether strain structure is possible and, if so, to identify the factors that drive its emergence. Our

model results demonstrated that inter-strain competition leads to a stratification of the different strain types due to the effects of competitive exclusion at a particular epitope or antigenic locus. Coupled with the effects of immunogenicity and the effects of host immunity targeting distinct viral antigens (explored in terms of epitope specificity in Theme II, but here encapsulating immunogenicity with respect to variant specificity), these factors characterised the establishment of one of three distinct types of viral strain structure: (i) no strain structure (NSS), defined by the stable co-existence of all possible strain types; (ii) discrete strain structure (DSS), representing a stable set of strains that do not share overlapping antigenic variants at either epitope; and (iii) partial discrete strain structure (PDSS), describing an intermediate stable state in between the previous two strain structures that consists of an overlapping subset of some, but not all, strains. We pose our results in the context of HTLV-I-associated disease and draw a comparison with clinical observations. In particular, our multi-locus HTLV-I model can provide a consistent mathematical framework to evaluate the experimental hypothesis that a greater degree of viral homogeneity is positively correlated with the presence and severity of HTLV-I-linked pathology (Gillet et al., 2011; Niewiesk et al., 1994), as well as raising plausible implications for the diagnosis and treatment of associated disease, thereby laying the groundwork for further experimental and theoretical study.

7.2 Extensions and Future Directions

The conclusions from our mathematical modelling investigations have helped us gain a more thorough understanding of the host-virus landscape in the context of chronic HTLV-I infection. Yet, there remain many unanswered questions relating to viral establishment, persistence, and pathogenesis. The work presented in this thesis is just another piece of a larger puzzle and there is certainly much more work that can be done, both from biological as well as mathematical standpoints. In particular, in the following we list several avenues of exploration stimulated from this work which we would like to pursue in the future.

- ▷ **Innate immunity and the humoral branch of adaptive immunity:** What effect do inhibitory killer-cell immunoglobulin-like receptors (KIRs) and antibodies, respectively, have on the host-virus interaction?
- ▷ **Incorporating variable degrees of cross-protection by host immunity:** How can we characterise the spectrum of cross-reactivity in the host-virus interaction, ranging from fully cross-reactive (as in our multi-locus model) to fully strain-specific?

- ▷ **Options for vaccine design:** What is the best way to augment HTLV-I-specific cellular immune responses to achieve a desired outcome, for instance, a reduced proviral load or a reduced activated proportion of proviral cells?
- ▷ **Measuring viral diversity:** How can we quantify viral diversity so that we can use such measures to aid us in identifying patterns that indicate effective control of viral burden or detriment?
- ▷ **Dynamics of T-cell activation:** Because the extent of proviral T-cell activation occurs on a spectrum, how does immune pressure select for the right level of activation and subsequent viral expression within the HTLV-I-infected target cell population?
- ▷ **Extending the multi-locus HTLV-I model to incorporate an arbitrary number of epitopes:** Can we unify the multiple epitope model with the multi-locus model to create an even more general model of HTLV-I that encapsulates antigenic variability in more than two viral epitopes?
- ▷ **Reduction of dimensionality for mathematical models of HTLV-I:** Is it possible to construct a lower-dimensional model representation that can capture the fundamental dynamical behaviour of the original system? If so, such a representation could lead to a reduction in computational and other resource expenditures.
- ▷ **Identifiability and parameter estimation from data:** Can we make use of our mathematical models of HTLV-I in conjunction with more extensive data sets to identify and estimate parameters of interest which may not be easily obtained experimentally?

7.3 Concluding Remarks

In this thesis, our overall aim has been to develop a rigorous theoretical framework to elucidate the underlying mechanisms responsible for infection, persistence, and pathogenesis in relation to the human retrovirus HTLV-I. We approach these issues using mathematical modelling, which allows us to analyse different aspects of the complex dynamics underlying the host-virus interface. Importantly, the basis of our model representations must rest on a solid mathematical foundation that is both consistent and biologically meaningful. Mathematical modelling can break apart the complexities inherent in within-host immunological interactions and allow us to focus on particular aspects to clarify the role of immune selection on HTLV-I infection.

With the rise in interdisciplinary collaboration and communication, it is becoming increasing clear that mathematical modelling, in conjunction with experimental data, can be utilised as an additional tool for developing and testing theoretical hypotheses that can explain and address the idiosyncrasies observed in chronic pathogenic infections such as HTLV-I and, eventually, may be able to offer insights that guide experimental design and inform clinical work. Further work on both the mathematical and biological forefronts, such as those touched upon briefly in the preceding section, is required to advance and refine the development of these mathematical models representing the HTLV-I system and, ultimately, to validate and evaluate their potential impact on an experimentally or clinically relevant scale.

Appendix A

Additional Derivations and Proofs

A.1 Derivation of the (Dimensional) Basic Reproduction Number R_0

Here we derive the *basic reproduction number for viral infection* (also commonly referred to as the *basic reproductive ratio*), defined by Equation (3.2), for the dimensional baseline model (3.1) using three different methods. The infection-free equilibrium for model (3.1) is given by $P_0 = (x_H, 0, 0, 0)$, where $x_H = \lambda/\mu_1$.

A.1.1 Method I. Using the Next Generation Operator

This method involves deriving the basic reproduction number for viral infection, R_0, using the next generation operator (Diekmann et al., 1990). View the Jacobian matrix at the infection-free equilibrium $J(P_0)$ as a block matrix

$$J(P_0) = \left[\begin{array}{c|cc|c} -\mu_1 & 0 & -\beta x_H & 0 \\ \hline 0 & -\tau-\mu_2 & \beta x_H + r & 0 \\ 0 & \tau & -\mu_3 & 0 \\ \hline 0 & 0 & \nu & -\mu_4 \end{array}\right]. \tag{A.1}$$

The upper-left and lower-right blocks each have a negative eigenvalue, which is given by their respective entries. Write the middle block in the following way:

$$\begin{aligned} J_{2\times 2}(P_0) &= \begin{bmatrix} -\tau-\mu_2 & \beta x_H + r \\ \tau & -\mu_3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \beta x_H + r \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \tau+\mu_2 & 0 \\ -\tau & \mu_3 \end{bmatrix}. \end{aligned}$$

Let

$$F = \begin{bmatrix} 0 & \beta x_H + r \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \tau+\mu_2 & 0 \\ -\tau & \mu_3 \end{bmatrix},$$

so that

$$V^{-1} = \frac{1}{\mu_3(\tau+\mu_2)} \begin{bmatrix} \mu_3 & 0 \\ \tau & \tau+\mu_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau+\mu_2} & 0 \\ \frac{\tau}{\mu_3(\tau+\mu_2)} & \frac{1}{\mu_3} \end{bmatrix}.$$

Then,

$$FV^{-1} = \begin{bmatrix} \frac{\tau}{\mu_3(\tau+\mu_2)}\big(\beta x_H + r\big) & \frac{1}{\mu_3}\big(\beta x_H + r\big) \\ 0 & 0 \end{bmatrix}.$$

The spectral radius of FV^{-1} is precisely R_0:

$$R_0 = \rho\big(FV^{-1}\big) = \frac{\tau}{\mu_3(\tau+\mu_2)}\Big(\beta x_H + r\Big).$$

A.1.2 Method II. Using the Stability Condition for P_0

Here we use the condition that the infection-free equilibrium P_0 is linearly unstable if and only if $R_0 > 1$ in order to derive an appropriate expression for R_0. The full proof of the stability of P_0 for the baseline model (3.1) is shown in Subsection 3.4.5. From the block matrix form of $J(P_0)$ shown in (A.1), we observe that since the upper-left and lower-right blocks each have a negative eigenvalue, the linear stability of $J(P_0)$ is determined by the eigenvalues of the middle 2×2 block of $J(P_0)$,

$$J_{2\times 2}(P_0) = \begin{bmatrix} -\tau - \mu_2 & \beta x_H + r \\ \tau & -\mu_3 \end{bmatrix}.$$

Using the well-known Routh-Hurwitz criterion for 2×2 matrices, since $\operatorname{tr}\big(J_{2\times 2}(P_0)\big) < 0$, it follows that $J_{2\times 2}(P_0)$ is linearly stable if and only if $\det\big(J_{2\times 2}(P_0)\big) > 0$. Hence, for instability, we require

$$\begin{aligned} & \det\big(J_{2\times 2}(P_0)\big) = \mu_3(\tau+\mu_2) - \tau\Big(\beta x_H + r\Big) < 0 \\ \iff & \frac{\tau}{\mu_3(\tau+\mu_2)}\Big(\beta x_H + r\Big) > 0 \\ \iff & R_0 > 1. \end{aligned}$$

Setting $R_0 = \dfrac{\tau}{\mu_3(\tau+\mu_2)}\Big(\beta x_H + r\Big)$ yields the expression for the basic reproduction number.

A.1.3 Method III. Using the Biological Interpretation

Biologically, the basic reproduction number may be expressed in the following way:

$$R_0 = I_T \cdot C_e = I_T \cdot C \cdot e,$$

where I_T is the mean infectious period and C_e, the infection rate or effective contact rate, is the product of two quantities, the contact rate C, and the effective rate of transmissibility of infected cells e.

In model (3.1), the rate at which a cell leaves the actively infected cell compartment is μ_3. One may think of the mean infectious period (i.e. the average duration in the y

compartment) as the reciprocal of this rate, so $I_T = 1/\mu_3$. It is observed that a fraction $\tau/(\tau+\mu_2)$ goes from the u-compartment to the y-compartment, representing the proportion of latently infected cells that begin to express the viral Tax protein, becoming actively infected and thus able to propagate the infection, so $e = \tau/(\tau + \mu_2)$. Lastly, the contact rate (which is a somewhat misleading name as it involves a rate due to mitosis) is given by the sum of the contribution of proviral cells from both infectious and mitotic transmission routes; that is, $C = \beta x_H + r$.

Therefore, we have derived

$$R_0 = I_T \cdot C \cdot e = \frac{\tau}{\mu_3(\tau + \mu_2)}\Big(\beta x_H + r\Big).$$

R_0 may be thought of as representing the average number of secondary infected cells resulting from a single actively infected cell introduced into a completely susceptible population of healthy cells over its lifetime (the mean infectious period).

A.2 Proofs for the Tax/HBZ Model (4.1)

A.2.1 Derivation of Expression for Tax-Specific CTL Dominance in Subsection 4.3.1

Suppose that $R_0 > 1$. Then by Theorem 4.2, there exists a globally asymptotically stable chronic infection steady state $P^* = (x^*, u^*, y^*, z^*, s^*)$, whose components satisfy the following equilibrium equations:

$$0 = \lambda - \beta x^* y^* - \mu_1 x^*, \tag{A.2a}$$

$$0 = \beta x^* y^* + r y^* - (\tau + \mu_2) u^*, \tag{A.2b}$$

$$0 = \tau u^* - \gamma y^* z^* - \gamma' y^* s^* - \mu_3 y^*, \tag{A.2c}$$

$$0 = \delta \nu y^* - \mu_4 z^*, \tag{A.2d}$$

$$0 = (1 - \delta)\nu' y^* - \mu_5 s^*. \tag{A.2e}$$

Equations (A.2a) to (A.2e) allow us to express x^*, u^*, z^*, and s^* in terms of y^*. Namely,

$$x^* = \frac{\lambda}{\beta y^* + \mu_1}, \quad u^* = \frac{1}{\tau}\left(\delta\gamma\phi_{\text{Tax}} y^* + (1-\delta)\gamma'\phi_{\text{HBZ}} y^* + \mu_3\right) y^*, \tag{A.3}$$

and

$$z^* = \delta\phi_{\text{Tax}} y^*, \quad s^* = (1-\delta)\phi_{\text{HBZ}} y^*, \tag{A.4}$$

where $\phi_{\text{Tax}} = \nu/\mu_4$, $\phi_{\text{HBZ}} = \nu'/\mu_5$, and y^* is a root of the following function $F(y)$:

$$F(y) = \frac{\tau}{\mu_3(\tau + \mu_2)}\left(\frac{\beta\lambda}{\beta y + \mu_1} + r\right) - \left(\frac{\delta\gamma}{\mu_3}\phi_{\text{Tax}} + (1-\delta)\frac{\gamma'}{\mu_3}\phi_{\text{HBZ}}\right) y - 1. \tag{A.5}$$

According to Equation (A.4), the relative abundance of Tax-specific CTLs to all CTLs at the chronic infection equilibrium can be described by

$$\frac{z^*}{z^*+s^*} = \frac{\delta\phi_{\text{Tax}}}{\delta\phi_{\text{Tax}} + (1-\delta)\phi_{\text{HBZ}}}. \tag{A.6}$$

We observe that the expression in Equation (A.6) is independent of the rates of lysis for either CTL type.

Next, we determine conditions on immunogenicity (δ) resulting in dominance of Tax-specific CTLs, defined as comprising greater than 50% of the total CTL population. Such a scenario occurs when

$$\frac{z^*}{z^*+s^*} > \frac{1}{2}.$$

That is, when

$$\delta > \frac{1}{\phi_{\text{Tax}}/\phi_{\text{HBZ}} + 1} =: \tilde{\delta}. \tag{A.7}$$

A.2.2 Proof of Theorem 4.3

Proof of Theorem 4.3. At the unique chronic infection equilibrium P^*, the equation $F(y^*) = 0$ holds, where $F(y)$ is the function defined in Equation (A.5). We differentiate this equation implicitly with respect to the parameter δ, which yields the following:

$$\frac{\mathrm{d}y^*}{\mathrm{d}\delta} = -A(\delta)\left(\frac{\gamma\nu}{\mu_4} - \frac{\gamma'\nu'}{\mu_5}\right)y^*, \tag{A.8}$$

where

$$A(\delta) = \left(\frac{\tau}{\tau+\mu_2}\frac{\beta^2\lambda}{(\beta y^*+\mu_1)^2} + \frac{\delta\gamma\nu}{\mu_4} + (1-\delta)\frac{\gamma'\nu'}{\mu_5}\right)^{-1} > 0 \text{ for all } \delta \in [0,1].$$

Consider the effective ratio of Tax-specific to HBZ-specific CTLs given by $\eta = \big(\gamma\phi_{\text{Tax}}\big)/\big(\gamma'\phi_{\text{HBZ}}\big)$. Then,

$$\frac{\mathrm{d}y^*}{\mathrm{d}\delta} \begin{cases} > 0 & \text{if } \eta < 1, \\ = 0 & \text{if } \eta = 1, \\ < 0 & \text{if } \eta > 1. \end{cases} \tag{A.9}$$

The proviral load at equilibrium is given by the expression $v^* = (u^*+y^*)/(x^*+u^*+y^*)$. To demonstrate that v^* is positively correlated with Tax immunogenicity δ, we compute

$$\frac{\mathrm{d}}{\mathrm{d}y^*}(u^*+y^*) = \frac{1}{\tau}\left(2\frac{\delta\gamma\nu}{\mu_4}y^* + 2(1-\delta)\frac{\gamma'\nu'}{\mu_5}y^* + \tau + \mu_3\right) > 0,$$

and

$$\frac{\mathrm{d}x^*}{\mathrm{d}y^*} = -\frac{\beta\lambda}{(\beta y^*+\mu_1)^2} = -\frac{\beta}{\beta y^*+\mu_1}x^* < 0.$$

Then,

$$\begin{aligned}
\frac{\mathrm{d}v^*}{\mathrm{d}y^*} &= \frac{\mathrm{d}}{\mathrm{d}y^*}\left(\frac{u^*+y^*}{x^*+u^*+y^*}\right) \\
&= \frac{(x^*+u^*+y^*)\dfrac{\mathrm{d}}{\mathrm{d}y^*}(u^*+y^*) - (u^*+y^*)\dfrac{\mathrm{d}}{\mathrm{d}y^*}(x^*+u^*+y^*)}{(x^*+u^*+y^*)^2} \\
&= \frac{1}{(x^*+u^*+y^*)^2}\left(x^*\frac{\mathrm{d}}{\mathrm{d}y^*}(u^*+y^*) - (u^*+y^*)\frac{\mathrm{d}x^*}{\mathrm{d}y^*}\right) \\
&= \frac{x^*}{(x^*+u^*+y^*)^2}\left[\frac{1}{\tau}\left(2\frac{\delta\gamma\nu}{\mu_4}y^* + 2(1-\delta)\frac{\gamma'\nu'}{\mu_5}y^* + \tau + \mu_3\right) + \frac{\beta(u^*+y^*)}{(\beta y^*+\mu_1)}\right] \\
&> 0.
\end{aligned}$$

Thus, the equilibrium proviral load v^* is positively correlated with the abundance of active proviral cells y^*. Since

$$\frac{\mathrm{d}v^*}{\mathrm{d}\delta} = \frac{\mathrm{d}v^*}{\mathrm{d}y^*}\frac{\mathrm{d}y^*}{\mathrm{d}\delta},$$

we therefore conclude that

$$\frac{\mathrm{d}v^*}{\mathrm{d}\delta}\begin{cases} > 0 & \text{if } \eta < 1, \\ = 0 & \text{if } \eta = 1, \\ < 0 & \text{if } \eta > 1. \end{cases} \tag{A.10}$$

This completes the proof of the theorem. □

A.3 Proofs for the General Multiple Epitope Model (5.1)

A.3.1 Biologically Realistic Region for the Multiple Epitope Model (5.1)

Proof of Theorem 5.1. Let $p \geq 1$. From the model equations, we notice that for any set of non-negative initial conditions, $x(t), u(t), y(t), z_k(t)$ remain non-negative for all $t \geq 0$ and $k = 1, \ldots, p$. It is clear that

$$\frac{\mathrm{d}x}{\mathrm{d}t} \leq \lambda - \mu_1 x,$$

and this implies that

$$\limsup_{t\to\infty} x(t) \leq \frac{\lambda}{\mu_1}.$$

Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$ and suppose that $r < \mu$. Adding the first three equations of model (5.1) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(x+u+y) \leq \lambda - (\mu - r)(x+u+y), \quad \text{where } \mu = \min\{\mu_1, \mu_2, \mu_3\},$$

which implies that $\limsup_{t\to\infty}(x+u+y)(t) \leq \lambda/(\mu-r)$. The assumption that $r < \mu$ ensures that the quantity on the right-hand side of the inequality is always positive. For the last inequality, let $\big(x(t), u(t), y(t), \vec{z}(t)\big)$ be a solution of model (5.1) subject to the condition $x(0) + u(0) + y(0) \leq \lambda/(\mu - r)$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=1}^{p} z_k\right) = \sum_{k=1}^{p} \frac{\mathrm{d}z_k}{\mathrm{d}t} = \frac{\sum_{k=1}^{p} \epsilon_k \nu_k}{\sum_{i=1}^{p} \epsilon_i} y - \sum_{k=1}^{p} \mu_{4,k} z_k \leq \tilde{\nu} y - \tilde{\mu} \sum_{k=1}^{p} z_k,$$

where $\tilde{\nu} = \max\limits_{k=1,\ldots,p} \nu_k$ and $\tilde{\mu} = \min\limits_{k=1,\ldots,p} \mu_{4,k}$, so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=1}^{p} z_k\right) \leq \frac{\lambda\tilde{\nu}}{\mu - r} - \tilde{\mu} \sum_{k=1}^{p} z_k \quad \Longrightarrow \quad \limsup_{t\to\infty} \sum_{k=1}^{p} z_k \leq \frac{\lambda\nu}{\tilde{\mu}(\mu - r)}.$$

Therefore, we consider the closed, bounded region

$$\Gamma_p := \left\{ (x, u, y, z_k) \in \mathbb{R}_+^{p+3} : x \leq \frac{\lambda}{\mu_1},\ x + u + y \leq \frac{\lambda}{\mu - r},\ \sum_{k=1}^{p} z_k \leq \frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu - r)} \right\}.$$

It is a straight-forward exercise to verify that Γ_p is positively invariant in $\mathbb{R}^{p+3}$ and hence model (5.1) is well-posed. □

A.3.2 Proof of Theorem 5.2. Part 1: Existence of Equilibria

The first proposition shows that the basic reproduction number for viral infection, given by

$$R_0 = \frac{\tau}{\mu_3(\tau + \mu_2)}\big(\beta x_H + r\big), \quad \text{where } x_H = \frac{\lambda}{\mu_1},$$

as in Equation (3.2), determines the existence of equilibria in $\bar{\Gamma_p}$.

Proposition A.1 (Existence of Equilibria)**.**

1) The infection-free equilibrium $P_0 = (x_H, 0, 0, \vec{0})$, where $x_H = \lambda/\mu_1$ and $\vec{0} \in \mathbb{R}^p$ is the p-dimensional null vector, always exists in $\bar{\Gamma_p}$. Moreover, if $0 < R_0 < 1$, then P_0 is the only equilibrium in $\bar{\Gamma_p}$;

2) If $R_0 > 1$, there exist exactly two equilibria, P_0 on $\partial\Gamma_p$ and a unique chronic infection equilibrium $P^* = (x^*, u^*, y^*, \vec{z}^*)$ in $\mathring{\Gamma}_p$, where $\vec{z}^* = (z_1^*, \ldots, z_p^*)$.

Proof. Equilibria occur when

$$0 = \lambda - \beta x y - \mu_1 x, \tag{A.11a}$$

$$0 = \beta x y + r y - (\tau + \mu_2) u, \tag{A.11b}$$

$$0 = \tau u - \gamma y z - \gamma' y s - \mu_3 y, \tag{A.11c}$$

$$0 = \frac{\varepsilon_k}{\sum_{i=1}^{p} \varepsilon_i} \nu_k y - \mu_{4,k} z_k, \quad \text{for each} \quad k = 1, \ldots, p. \tag{A.11d}$$

Observe that since x, u, y, z_k are restricted to be non-negative in the biologically feasible region Γ_p, then from the equilibrium equations (A.11c) and (A.11d), at equilibrium $u = 0 \iff y = 0 \iff z_k = 0$ for all $k = 1, \ldots, p$. We will demonstrate that there are only two possible non-negative equilibria in our model (5.1): P_0 representing the infection-free state, and P^* representing chronic infection.

We first determine the existence of the *infection-free* equilibrium $P_0 = (x_H, 0, 0, \vec{0})$, where $\vec{0} \in \mathbb{R}^p$ and $x_H = \frac{\lambda}{\mu_1}$ is the level of target cells in the absence of an infection and coincides precisely with the normal CD4$^+$ helper T-cell count in a healthy individual.

Next, we search for steady state solutions of the form $P^* = (x^*, u^*, y^*, \vec{z}^*)$, where $\vec{z}^* = (z_1^*, \ldots, z_p^*)$ and $x^*, u^*, y^*, z_k^* > 0$ for all $k = 1, \ldots, p$. We will refer to such a steady state, if it exists, as a *chronic infection* equilibrium. Equations (A.11a) to (A.11d) allow us to express x^*, u^*, and z_k^* for $k = 1, \ldots, p$ in terms of y^*. Namely, the components of P^* must satisfy

$$x^* = \frac{\lambda}{\beta y^* + \mu_1}, \quad u^* = \frac{1}{\tau \sum_{k=1}^p \varepsilon_i} \left(\sum_{k=1}^p \varepsilon_k \frac{\gamma_k \nu_k}{\mu_{4,k}} \right) y^{*2} + \frac{\mu_3}{\tau} y^*, \quad z_k^* = \frac{\varepsilon_k}{\sum_{i=1}^p \varepsilon_i} \frac{\nu_k}{\mu_{4,k}} y^*, \tag{A.12}$$

where y^* is a root of the following function $F_p(y)$:

$$F_p(y) = \frac{\tau}{\mu_3(\tau + \mu_2)} \left(\frac{\beta\lambda}{\beta y + \mu_1} + r \right) - \frac{1}{\mu_3 \sum_{i=1}^p \varepsilon_i} \sum_{k=1}^p \frac{\varepsilon_k \gamma_k \nu_k}{\mu_{4,k}} y - 1. \tag{A.13}$$

It is straight-forward to show that $F_p(y)$ is a monotonically decreasing function of y $\big($i.e. $F_p'(y) < 0\big)$ and $\lim_{y\to\infty} F_p(y) < 0$. Hence, since

$$F_p(0) = \big[R_0 - 1\big],$$

the existence of a unique positive root y^* of $F_p(y) = 0$ is equivalent to the condition that $R_0 > 1$, thus completing the proof. □

A.3.3 Proof of Theorem 5.2. Part 2: Stability of P_0

The Jacobian matrix for the multiple epitope model (5.1) is a $(p+3)$-dimensional matrix

$$J(x, u, y, \vec{z}) = \begin{array}{c} \\ k=1 \\ \\ \\ \\ \\ \\ k=p \end{array} \begin{array}{c} \begin{array}{ccccccccc} & & & & k=1 & & & & k=p \end{array} \\ \begin{bmatrix} -\beta y - \mu_1 & 0 & -\beta x & 0 & \cdots & \cdots & \cdots & 0 \\ \beta y & -\tau - \mu_2 & \beta x + r & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \tau & -\sum_{i=1}^p \gamma_i z_i - \mu_3 & -\gamma_1 y & \cdots & \cdots & \cdots & -\gamma_p y \\ 0 & 0 & \frac{\varepsilon_1}{\sum_{i=1}^p \varepsilon_i} \nu_1 & -\mu_{4,1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \frac{\varepsilon_1}{\sum_{i=1}^p \varepsilon_i} \nu_p & 0 & \cdots & \cdots & 0 & -\mu_{4,p} \end{bmatrix} \end{array},$$

(A.14)

which can be written in block matrix form as

$$J(x,u,y,\vec{z}) = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right],$$

where

$$A_{11} = \underbrace{\begin{bmatrix} -\beta y - \mu_1 & 0 & -\beta x \\ \beta y & -\tau - \mu_2 & \beta x + r \\ 0 & \tau & -\sum_{k=1}^{p} \gamma_k z_k - \mu_3 \end{bmatrix}}_{3\times 3 \text{ matrix}}, \quad A_{12} = \underbrace{\begin{bmatrix} \vec{0}^T \\ \vec{0}^T \\ \left[(-\gamma_k y)_{k=1}^{p}\right]^T \end{bmatrix}}_{3\times p \text{ matrix}},$$

$$A_{21} = \underbrace{\begin{bmatrix} \vec{0} & \vec{0} & \left[\frac{\varepsilon_k}{\sum_{i=1}^{p}\varepsilon_i}\nu_k\right]_{k=1}^{p} \end{bmatrix}}_{p\times 3 \text{ matrix}}, \quad A_{22} = \underbrace{\begin{bmatrix} \underset{k=1,\ldots,p}{\text{diag}}\ (-\mu_{4,k}) \end{bmatrix}}_{p\times p \text{ matrix}}.$$

In the above, $\vec{0} = (0,\ldots,0) \in \mathbb{R}^p$ denotes the p-dimensional null vector. The local stability of P_0 is given by the following.

Proposition A.2 (Local Stability of P_0)**.**

i) When $R_0 < 1$, the infection-free equilibrium $P_0 = (x_H, 0, 0, \vec{0})$, where $x_H = \lambda/\mu_1$, is always locally asymptotically stable in the feasible region Γ_p.

ii) When $R_0 > 1$, P_0 is unstable. Specifically, P_0 is a saddle with $\dim W^s_{loc}(P_0) = (p+2)$ and $\dim W^u_{loc}(P_0) = 1$, where $W^s_{loc}(P_0)$, $W^u_{loc}(P_0)$ denote the local stable and unstable manifolds of P_0, respectively.

Proof. Establishing the local stability of the infection-free equilibrium P_0 is equivalent to determining the signs of the real parts of the eigenvalues of $J(P_0)$, the Jacobian matrix at P_0. To do this, we first calculate the characteristic polynomial for $J(P_0)$, defined by

$$\chi_0(\zeta) = \det\Big(\zeta I_{(p+3)\times(p+3)} - J(P_0)\Big),$$

where $I_{(p+3)\times(p+3)}$ denotes the $(p+3)\times(p+3)$ identity matrix. We evaluate

$$A_{11}\Big|_{P_0} = \begin{bmatrix} -\mu_1 & 0 & -\beta x_H \\ 0 & -\tau - \mu_2 & \beta x_H + r \\ 0 & \tau & -\mu_3 \end{bmatrix}, \quad A_{12}\Big|_{P_0} = \begin{bmatrix} \vec{0}^T \\ \vec{0}^T \\ \vec{0}^T \end{bmatrix},$$

$$A_{21}\Big|_{P_0} = \begin{bmatrix} \vec{0} & \vec{0} & \left[\frac{\varepsilon_k}{\sum_{i=1}^{p}\varepsilon_i}\nu_k\right]_{k=1}^{p} \end{bmatrix}, \quad A_{22}\Big|_{P_0} = \begin{bmatrix} \underset{k=1,\ldots,p}{\text{diag}}\ (-\mu_{4,k}) \end{bmatrix}.$$

Let $\overline{\mathbf{0}}_{3\times p}$ denote the $(3\times p)$ zero matrix. Since we notice that $J(P_0)$ is of a lower (block) triangular form

$$J(P_0) = \left[\begin{array}{c|c} * & \overline{\mathbf{0}}_{3\times p} \\ \hline * & * \end{array}\right],$$

this is equivalent simply to the product of the determinants of the diagonal block matrices

$$\Big(\zeta I_{3\times 3} - A_{11}\Big|_{P_0}\Big) \quad \text{and} \quad \Big(\zeta I_{p\times p} - A_{22}\Big|_{P_0}\Big),$$

where $I_{3\times 3}$ and $I_{p\times p}$ denote the appropriately sized identity matrices. Thus, the characteristic polynomial is

$$\chi_0(\zeta) = (\zeta+\mu_1)\left[\prod_{k=1}^{p}(\zeta+\mu_{4,k})\right]\Big[\zeta^2 + (\tau+\mu_2+\mu_3)\zeta - \mu_3(\tau+\mu_2)\big[R_0-1\big]\Big],$$

where R_0 is the basic reproduction number for viral infection defined in Equation (3.2). This is a straight-forward extension of the method used to calculate the characteristic polynomial for the baseline model (3.1), applied here to cover the situation pertaining to any arbitrary number p of epitopes as in model (5.1) (c.f. the proof of Proposition 3.2). The roots of $\chi_0(\zeta)$ are given by $\zeta_i \in \mathbb{C}$ such that $\chi_0(\zeta_i) = 0$:

$$\zeta_k = -\mu_{4,k} \quad \text{for} \quad k = 1,\ldots,p,$$

$$\zeta_{p+1} = -\mu_1,$$

and

$$\zeta_{p+2,p+3} = -\frac{1}{2}(\tau+\mu_2+\mu_3) \pm \frac{1}{2}\sqrt{(\tau+\mu_2+\mu_3)^2 + 4\mu_3(\tau+\mu_2)[R_0-1]}.$$

Clearly, $\zeta_1,\ldots,\zeta_{p+1} < 0$. Moreover, if $R_0 < 1$, then $\mathrm{Re}(\zeta_{p+2})$ and $\mathrm{Re}(\zeta_{p+3}) < 0$. Hence P_0 is either a stable node or a stable spiral. However, if $R_0 > 1$, it is readily seen that $\mathrm{Re}(\zeta_{p+2}) > 0$, so that P_0 is unstable. Specifically, P_0 is a saddle whose local stable manifold $W^s_{loc}(P_0)$ has dimension $(p+2)$ and local unstable manifold $W^u_{loc}(P_0)$ has dimension 1. If $R_0 = 1$, then $\zeta_{p+2} = 0$ is a zero eigenvalue of $J(P_0)$ and no immediate conclusions about the local stability of P_0 may be inferred. The preceding argument establishes the basic reproduction number for viral infection R_0 as a threshold parameter that characterises the local stability of the infection-free equilibrium P_0. □

In fact, a stronger statement can be made about the stability of P_0 when $R_0 < 1$, namely that it is globally asymptotically stable in the feasible region Γ_p.

Proposition A.3 (Global Stability of P_0)**.** *When $R_0 < 1$, the infection-free equilibrium P_0 is globally asymptotically stable in the feasible region Γ_p.*

Proof. Assume that $R_0 < 1$. Then by Proposition A.1, the infection-free steady state P_0 is the only equilibrium point in $\bar{\Gamma}_p$. Note that $x \leq x_H$ for all $x \in \Gamma_p$. In the following, we will assume that $\epsilon_k > 0$ for all $k = 1,\ldots,p$. The case when $\epsilon_{k_0} = 0$ for some $k_0 \in \{1,\ldots,p\}$ will be discussed at the end of the proof.

Consider

$$L_p(x,u,y,z_1,\ldots,z_p)=\left(x-x_H-x_H\log\frac{x}{x_H}\right)+u+\frac{\tau+\mu_2}{\tau}y+\frac{(\tau+\mu_2)\sum_{i=1}^{p}\varepsilon_i}{2\tau}\sum_{k=1}^{p}\frac{\gamma_k}{\varepsilon_k\nu_k}z_k^2.$$

The function $L_p:\mathbb{R}_+^{p+3}\to\mathbb{R}$ is positive definite since

(i) $L_p(x_H,0,0,\vec{0})=0$, i.e. $L_p(P_0)=0$; and

(ii) $L_p(x,u,y,\vec{z})>0$, for all $(x,u,y,\vec{z})\in\mathbb{R}_+^{p+3},(x,u,y,\vec{z})\neq(x_H,0,0,\vec{0})$.

The Lyapunov derivative of L_p is

$$\begin{aligned}\frac{\mathrm{d}L_p}{\mathrm{d}t}&=\left(1-\frac{x_H}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\mathrm{d}u}{\mathrm{d}t}+\frac{\tau+\mu_2}{\tau}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{(\tau+\mu_2)\sum_{i=1}^{p}\varepsilon_i}{\tau}\sum_{k=1}^{p}\frac{\gamma_k}{\varepsilon_k\nu_k}z_k\frac{\mathrm{d}z}{\mathrm{d}t}\\&=\left(1-\frac{x_H}{x}\right)(\lambda-\mu_1x)-\beta xy\left(1-\frac{x_H}{x}\right)+\Big(\beta xy+ry-(\tau+\mu_2)u\Big)\\&\quad+\frac{\tau+\mu_2}{\tau}\left(\tau u-y\sum_{k=1}^{p}\gamma_kz_k-\mu_3y\right)\\&\quad+\frac{(\tau+\mu_2)\sum_{i=1}^{p}\varepsilon_i}{\tau}\sum_{k=1}^{p}\frac{\gamma_k}{\varepsilon_k\nu_k}z_k\left[\frac{\varepsilon_k}{\sum_{i=1}^{p}\varepsilon_i}\nu_ky-\mu_{4,k}z_k\right].\end{aligned}$$

Replacing $\lambda=\mu_1x_H$ and expanding terms yields

$$\begin{aligned}\frac{\mathrm{d}L_p}{\mathrm{d}t}&=-\frac{\mu_1}{x}(x-x_H)^2+\frac{\mu_3}{\tau}(\tau+\mu_2)\big[R_0-1\big]y-\frac{(\tau+\mu_2)\sum_{i=1}^{p}\varepsilon_i}{\tau}\sum_{k=1}^{p}\frac{\gamma_k\mu_{4,k}}{\varepsilon_k\nu_k}z_k^2\\&\leq 0,\quad\forall(x,u,y,\vec{z})\in\Gamma_p,\quad\text{since } R_0<1.\end{aligned}$$

Hence, L_p is an appropriate Lyapunov function and moreover,

$$\frac{\mathrm{d}L_p}{\mathrm{d}t}(x_H,0,0,\vec{0})=L_p(x_H,0,0,\vec{0})=0.$$

Since the feasible region Γ_p is compact and invariant with respect to the vector field of model (5.1), it then follows from LaSalle's invariance principle (LaSalle, 1976) that every trajectory starting in Γ_p approaches a set M_p, where M_p is the largest invariant subset of

$$I_p=\left\{(x,u,y,\vec{z})\in\mathbb{R}_+^{p+3}:\frac{\mathrm{d}L_p}{\mathrm{d}t}(x,u,y,\vec{z})=0\right\},$$

which, in the present case, is precisely

$$\{x\equiv x_H\}\cap\{y\equiv 0\}\cap\bigcap_{k=1}^{p}\{z_k\equiv 0\}.$$

To establish the global stability of P_0 when $R_0<1$, it remains to show that the set M_p consists solely of the equilibrium point P_0. Observe that in the set I_p, the variable u can take on any arbitrary value along the non-negative real line. Let $\omega_p(t)=$

$\big(x(t), u(t), y(t), z_1(t), \ldots, z_p(t)\big)$ be any trajectory of model (5.1) starting in I_p, so that $x(0) = x_H$, $y(0) = 0$, and $z_k(0) = 0$ for $k = 1, \ldots, p$. Note that the value of $u(0) = u_0$ is not yet known at this stage. Then, the behaviour of $\omega_p(t)$ at $t = 0$ is governed by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}u}{\mathrm{d}t} = -(\tau + \mu_2)u, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = \tau u, \quad \text{and} \quad \frac{\mathrm{d}z_k}{\mathrm{d}t} = 0 \quad \text{for } k = 1, \ldots, p.$$

Hence, $x(t) \equiv x_H$, $\vec{z}(t) \equiv \vec{0}$, and

$$u(t) = u(0)e^{-(\tau+\mu_2)t}.$$

The y-compartment varies according to $\mathrm{d}y/\mathrm{d}t = \tau u(0)$ at time $t = 0$. If $u(0) \neq 0$, then $\mathrm{d}y/\mathrm{d}t > 0$, implying that $y(t_0) \neq 0$ for some $t_0 \neq 0$; that is, the trajectory will leave the set I_p. For the solution $\omega_p(t)$ to remain in I_p for all t, we must have $u(0) = 0$ and therefore $u(t) \equiv 0$. Subsequently, solving the initial value problem $\mathrm{d}y/\mathrm{d}t = 0$, $y(0) = 0$ yields $y(t) \equiv 0$. Taken all together, the above argument means that the maximal invariant set in I_p is comprised only of the single point $P_0 = (x_H, 0, 0, \vec{0})$, i.e. $M_p = \{P_0\}$. □

Remark A.1 (Establishing the global stability of P_0 when ε_{k_0} takes a value of 0)**.** For simplicity, let $p = 2$ so that we are in the two-epitope case. The model can then be reduced to the Tax/HBZ model (4.1) by setting $\delta = \varepsilon_1/(\varepsilon_1 + \varepsilon_2)$. Assume that δ takes a value at one of the endpoints, i.e. $\delta = 0$ or $\delta = 1$. Then, either the z-equation (in the former case), or the s-equation (in the latter case) in model (4.1) would describe exponential decay of that particular variable. In both situations, the model dynamics would be further reduced to that of the baseline HTLV-I model (3.1). The corresponding Lyapunov functions take on a similar form as the function L above, with minor modifications depending on the precise choice of δ (i.e. either 0 or 1) as indicated in the following.

First, let us suppose that $\delta = 1$. Then, the equations for the Tax- and HBZ-specific CTL compartments in model (4.1) become $\mathrm{d}z/\mathrm{d}t = \nu y - \mu_4 z$ and $\mathrm{d}s/\mathrm{d}t = -\mu_5 z$, respectively. Consider

$$L_{\delta=1} = L_{\delta=1}(x, u, y, z, s) = \left(x - x_H - x_H \log \frac{x}{x_H}\right) + u + \frac{\tau + \mu_2}{\tau} y + \frac{\gamma(\tau + \mu_2)}{2\tau\nu} z^2 + s.$$

As in the general case above, it can be easily verified that the function $L_{\delta=1}$ is positive

definite. We compute the Lyapunov derivative of $L_{\delta=1}$ as follows:

$$\begin{aligned}
\frac{\mathrm{d}L_{\delta=1}}{\mathrm{d}t} &= \left(1-\frac{x_H}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\tau+\mu_2}{\tau}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\gamma(\tau+\mu_2)}{\delta\tau\nu}z\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\mathrm{d}s}{\mathrm{d}t} \\
&= \left(1-\frac{x_H}{x}\right)(\lambda-\mu_1 x) - \beta xy\left(1-\frac{x_H}{x}\right) + \Big(\beta xy + ry - (\tau+\mu_2)u\Big) \\
&\quad + \frac{\tau+\mu_2}{\tau}(\tau u - \gamma yz - \gamma' ys - \mu_3 y) + \frac{\gamma(\tau+\mu_2)}{\tau\nu}z(\nu y - \mu_4 z) + (-\mu_5 s) \\
&= -\frac{\mu_1}{x}(x-x_H)^2 + \frac{\mu_3}{\tau}(\tau+\mu_2)\big[R_0-1\big]y - \frac{\gamma\mu_4(\tau+\mu_2)}{\tau\nu}z^2 - \frac{\gamma'(\tau+\mu_2)}{\tau}ys \\
&\quad - \mu_5 s \\
&\leq 0, \quad \forall (x,u,y,z,s)\in\Gamma, \quad \text{since } R_0<1.
\end{aligned}$$

Therefore, $L_{\delta=1}$ is a suitable Lyapunov function and we can argue using LaSalle's invariance principle (LaSalle, 1976), as we have done in the general case, that all trajectories of model (4.1) for $\delta=1$ converge to the infection-free steady state P_0 when $R_0<1$.

Analogously, if $\delta=0$, let

$$L_{\delta=0} = L_{\delta=0}(x,u,y,z,s) = \left(x - x_H - x_H\log\frac{x}{x_H}\right) + u + \frac{\tau+\mu_2}{\tau}y + z + \frac{\gamma'(\tau+\mu_2)}{2\tau\nu'}s^2.$$

Then, it can be similarly shown that

$$\frac{\mathrm{d}L_{\delta=0}}{\mathrm{d}t} \leq 0, \quad \forall (x,u,y,z,s)\in\Gamma, \quad \text{when } R_0<1.$$

Hence, the function $L_{\delta=0}$ is an appropriate Lyapunov function that establishes the global convergence of P_0 for the situation when $\delta=0$.

A.3.4 Proof of Theorem 5.2. Part 3: Stability of P^*

The final part of the proof of Theorem 5.2 is to demonstrate the global stability of the chronic infection equilibrium P^* whenever it exists in Γ_p. We therefore establish the following.

Proposition A.4 (Global Stability of P^*)**.** *When $R_0>1$, the chronic-infection equilibrium P^* is globally asymptotically stable in $\mathring{\Gamma}_p$.*

Proof. Assume that $R_0>1$. Then by Proposition A.1, the chronic infection equilibrium P^* exists in the feasible region Γ_p. To prove the result, we construct an appropriate Lyapunov function $V_p : \mathbb{R}_+^{p+3} \to \mathbb{R}$ that will allow us to conclude the global asymptotic stability properties of P^* in $\mathring{\Gamma}_p$. As with the proof for the global stability of the infection-free steady state in Proposition A.3, in the following, we will assume that $\epsilon_k>0$ for all $k=1,\ldots,p$. We will discuss the special case when $\epsilon_{k_0}=0$ for some $k_0\in\{1,\ldots,p\}$ at the end of the proof.

Suppose that $\big(x(t), u(t), y(t), \vec{z}(t)\big)$ is a solution of model (5.1) and let

$$V_p = \left(x - x^* - x^* \log \frac{x}{x^*}\right) + \left(u - u^* - u^* \log \frac{u}{u^*}\right) + \frac{\tau + \mu_2}{\tau}\left(y - y^* - y^* \log \frac{y}{y^*}\right) + \frac{(\tau + \mu_2)\sum_{i=1}^{p} \varepsilon_i}{\tau} \sum_{k=1}^{p} \frac{\gamma_k}{\varepsilon_k \nu_k} z_k^* \left(z_k - z_k^* - z_k^* \log \frac{z_k}{z_k^*}\right).$$

The function $V_p = V_p(x, u, y, \vec{z})$ is positive definite with respect to the chronic infection equilibrium point $P^* = (x^*, u^*, y^*, \vec{z}^*)$ (refer to Lemma B.1). Taking the Lyapunov derivative along the solution $\big(x(t), u(t), y(t), \vec{z}(t)\big)$ yields

$$\begin{aligned}
\frac{\mathrm{d}V_p}{\mathrm{d}t} &= \left(1 - \frac{x^*}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \left(1 - \frac{u^*}{u}\right)\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\tau + \mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)\frac{\mathrm{d}y}{\mathrm{d}t} \\
&\quad + \frac{(\tau + \mu_2)\sum_{i=1}^{p} \varepsilon_i}{\tau} \sum_{k=1}^{p} \frac{\gamma_k}{\varepsilon_k \nu_k} z_k^* \left(1 - \frac{z_k^*}{z_k}\right)\frac{\mathrm{d}z_k}{\mathrm{d}t} \\
&= \left(1 - \frac{x^*}{x}\right)(\lambda - \beta x y - \mu_1 x) + \left(1 - \frac{u^*}{u}\right)\Big(\beta x y + r y - (\tau + \mu_2) u\Big) \\
&\quad + \frac{\tau + \mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)\left(\tau u - y \sum_{k=1}^{p} \gamma_k z_k - \mu_3 y\right) \\
&\quad + \frac{(\tau + \mu_2)\sum_{i=1}^{p} \varepsilon_i}{\tau} \sum_{k=1}^{p} \frac{\gamma_k}{\varepsilon_k \nu_k} z_k^* \left(1 - \frac{z_k^*}{z_k}\right)\left[\frac{\varepsilon_k}{\sum_{i=1}^{p} \varepsilon_i} \nu_k y - \mu_{4,k} z_k\right].
\end{aligned}$$

By making the appropriate substitutions using the equilibrium equations (A.11a)–(A.11d),

we find that

$$\begin{aligned}
\frac{\mathrm{d}V_p}{\mathrm{d}t} &= \left(1-\frac{x^*}{x}\right)(\mu_1 x^* - \mu_1 x) + \left(1-\frac{x^*}{x}\right)(\beta x^* y^* - \beta x y) \\
&\quad + \left(1-\frac{u^*}{u}\right)\left(\beta x y + r y - \left(\beta x^* y^* + r y^*\right)\frac{u}{u^*}\right) \\
&\quad + \frac{\beta x^* y^* + r y^*}{\tau u^*}\left(1-\frac{y^*}{y}\right)\left(\tau u - \tau u^* \frac{y}{y^*}\right) \\
&\quad + \frac{\tau+\mu_2}{\tau}\left(1-\frac{y^*}{y}\right)\left(y\sum_{k=1}^{p}\gamma_k z_k^* - y\sum_{k=1}^{p}\gamma_k z_k\right) \\
&\quad + \frac{(\tau+\mu_2)\sum_{i=1}^{p}\varepsilon_i}{\tau}\sum_{k=1}^{p}\frac{\gamma_k}{\varepsilon_k\nu_k}z_k^*\left(1-\frac{z_k^*}{z_k}\right)\left(\frac{\varepsilon_k\nu_k}{\sum_{i=1}^{p}\varepsilon_i}y - \frac{\varepsilon_k\nu_k}{\sum_{i=1}^{p}\varepsilon_i}y^*\frac{z_k}{z_k^*}\right) \\
&= \mu_1 x^*\left(2-\frac{x^*}{x}-\frac{x}{x^*}\right) + \beta x^* y^*\left(3-\frac{x^*}{x}-\frac{x}{x^*}\frac{u^*}{u}\frac{y}{y^*}-\frac{u}{u^*}\frac{y^*}{y}\right) \\
&\quad + r y^*\left(2-\frac{u^*}{u}\frac{y}{y^*}-\frac{u}{u^*}\frac{y^*}{y}\right) + \frac{(\tau+\mu_2)}{\tau}y\sum_{k=1}^{p}\gamma_k z_k^*\left[1-\frac{z_k}{z_k^*}-\frac{y^*}{y}+\frac{y^*}{y}\frac{z_k}{z_k^*}\right] \\
&\quad + \frac{(\tau+\mu_2)}{\tau}y\sum_{k=1}^{p}\gamma_k z_k^*\left[1-\frac{y^*}{y}\frac{z_k}{z_k^*}-\frac{z_k^*}{z_k}+\frac{y^*}{y}\right] \\
&= \mu_1 x^*\left(2-\frac{x^*}{x}-\frac{x}{x^*}\right) + \beta x^* y^*\left(3-\frac{x^*}{x}-\frac{x}{x^*}\frac{u^*}{u}\frac{y}{y^*}-\frac{u}{u^*}\frac{y^*}{y}\right) \\
&\quad + r y^*\left(2-\frac{u^*}{u}\frac{y}{y^*}-\frac{u}{u^*}\frac{y^*}{y}\right) + \frac{(\tau+\mu_2)}{\tau}y\sum_{k=1}^{p}\gamma_k z_k^*\left(2-\frac{z_k}{z_k^*}-\frac{z_k^*}{z_k}\right) \\
&\leq 0.
\end{aligned}$$

The final line follows from the inequalities $\sqrt{a_1 a_2} \leq (a_1+a_2)/2$ and $\sqrt[3]{a_1 a_2 a_3} \leq (a_1+a_2+a_3)/3$, where $a_1, a_2, a_3 > 0$; that is, the geometric mean of n positive numbers can be no greater than their arithmetic mean. Hence, the global stability of P^* has been established when $R_0 > 1$. □

Remark A.2 (Establishing the global stability of P^* when $\varepsilon_{k_0} = 0$)**.** To complete the proof, we must deal with the case when $\varepsilon_{k_0} = 0$ for some $k_0 \in \{1, \ldots, p\}$. For simplicity, let us consider the two-epitope case (i.e. $p = 2$), which is equivalent to the Tax/HBZ model (4.1), for example, by setting $\delta = \varepsilon_1/(\varepsilon_1+\varepsilon_2)$. First, suppose that $\delta = 1$. Then, the equations for the Tax- and HBZ-specific CTL compartments in model (4.1) become $\mathrm{d}z/\mathrm{d}t = \nu y - \mu_4 z$ and $\mathrm{d}s/\mathrm{d}t = -\mu_5 z$, respectively. We notice that the latter equation is decoupled from the remaining equations, and we can explicitly solve it to obtain

$$s(t) = s(0)\exp^{-\mu_5 t},$$

which implies that $s^* = 0$ (as $s^* = \lim_{t\to\infty} s(t)$). Define

$$V_{\delta=1} = \left(x - x^* - x^* \log \frac{x}{x^*}\right) + \left(u - u^* - u^* \log \frac{u}{u^*}\right) + \frac{\tau+\mu_2}{\tau}\left(y - y^* - y^* \log \frac{y}{y^*}\right) + \frac{\gamma}{\tau\nu}(\tau+\mu_2)z^*\left(z - z^* - z^* \log \frac{z}{z^*}\right) + \frac{\gamma'}{\tau\mu_5}(\tau+\mu_2)y^* s.$$

The function $V_{\delta=1} = V_{\delta=1}(x, u, y, z, s)$ is positive definite with respect to the chronic infection equilibrium point $P^* = (x^*, u^*, y^*, z^*, s^*)$. Calculating the Lyapunov derivative yields

$$\begin{aligned}
\frac{\mathrm{d}V_{\delta=1}}{\mathrm{d}t} &= \left(1 - \frac{x^*}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \left(1 - \frac{u^*}{u}\right)\frac{\mathrm{d}u}{\mathrm{d}t} + \frac{\tau+\mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)\frac{\mathrm{d}y}{\mathrm{d}t} \\
&\quad + \frac{\gamma}{\tau\nu}(\tau+\mu_2)z^*\left(1 - \frac{z^*}{z}\right)\frac{\mathrm{d}z}{\mathrm{d}t} + \frac{\gamma'}{\tau\mu_5}(\tau+\mu_2)y^*\frac{\mathrm{d}s}{\mathrm{d}t} \\
&= \left(1 - \frac{x^*}{x}\right)(\lambda - \beta xy - \mu_1 x) + \left(1 - \frac{u^*}{u}\right)\Big(\beta xy + ry - (\tau+\mu_2)u\Big) \\
&\quad + \frac{\tau+\mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)(\tau u - \gamma yz - \gamma' ys - \mu_3 y) \\
&\quad + \frac{\gamma}{\tau\nu}(\tau+\mu_2)z^*\left(1 - \frac{z^*}{z}\right)(\nu y - \mu_4 z) + \frac{\gamma'}{\tau\mu_5}(\tau+\mu_2)y^*(-\mu_5 s).
\end{aligned}$$

Using the equilibrium equations, we can make the appropriate substitutions and re-arrange the terms to obtain

$$\begin{aligned}
\frac{\mathrm{d}V_{\delta=1}}{\mathrm{d}t} &= \left(1 - \frac{x^*}{x}\right)(\mu_1 x^* - \mu_1 x) + \left(1 - \frac{x^*}{x}\right)(\beta x^* y^* - \beta xy) \\
&\quad + \left(1 - \frac{u^*}{u}\right)\left(\beta xy + ry - \left(\beta x^* y^* + ry^*\right)\frac{u}{u^*}\right) \\
&\quad + \frac{\beta x^* y^* + ry^*}{\tau u^*}\left(1 - \frac{y^*}{y}\right)\left(\tau u - \tau u^* \frac{y}{y^*}\right) \\
&\quad + \frac{\tau+\mu_2}{\tau}\left(1 - \frac{y^*}{y}\right)(\gamma y z^* - \gamma yz - \gamma' ys) \\
&\quad + \frac{\gamma}{\tau\nu}(\tau+\mu_2)z^*\left(1 - \frac{z^*}{z}\right)\left(\nu y - \nu y^* \frac{z}{z^*}\right) - \frac{\gamma'}{\tau}(\tau+\mu_2)y^* s \\
&= \mu_1 x^*\left(2 - \frac{x^*}{x} - \frac{x}{x^*}\right) + \beta x^* y^*\left(3 - \frac{x^*}{x} - \frac{x}{x^*}\frac{u^*}{u}\frac{y}{y^*} - \frac{u}{u^*}\frac{y^*}{y}\right) \\
&\quad + ry^*\left(2 - \frac{u^*}{u}\frac{y}{y^*} - \frac{u}{u^*}\frac{y^*}{y}\right) + \frac{\gamma}{\tau}(\tau+\mu_2)yz^*\left(2 - \frac{z}{z^*} - \frac{z^*}{z}\right) - \frac{\gamma'}{\tau}(\tau+\mu_2)ys \\
&\leq 0, \quad \forall (x, u, y, z, s) \in \mathring{\Gamma}.
\end{aligned}$$

As in the general case, the final line follows from the inequalities $\sqrt{a_1 a_2} \leq (a_1 + a_2)/2$ and $\sqrt[3]{a_1 a_2 a_3} \leq (a_1 + a_2 + a_3)/3$, where $a_1, a_2, a_3 > 0$. This shows that $V_{\delta=1}$ is a suitable Lyapunov function that establishes the global stability of P^* for model (4.1) in the case $\delta = 1$ when $R_0 > 1$.

Analogously, if $\delta = 0$, let

$$V_{\delta=0} = \left(x - x^* - x^* \log \frac{x}{x^*}\right) + \left(u - u^* - u^* \log \frac{u}{u^*}\right) + \frac{\tau + \mu_2}{\tau}\left(y - y^* - y^* \log \frac{y}{y^*}\right) + \frac{\gamma}{\tau\mu_4}(\tau + \mu_2)y^* z + \frac{\gamma'}{\tau\nu'}(\tau + \mu_2)s^*\left(s - s^* - s^* \log \frac{s}{s^*}\right).$$

Then, it can be similarly shown that

$$\frac{\mathrm{d}V_{\delta=0}}{\mathrm{d}t} \leq 0, \quad \forall (x, u, y, z, s) \in \mathring{\Gamma}.$$

Hence, the function $V_{\delta=0}$ is an appropriate Lyapunov function that establishes the global convergence of P^* for the Tax/HBZ model for $\delta = 0$ when $R_0 > 1$.

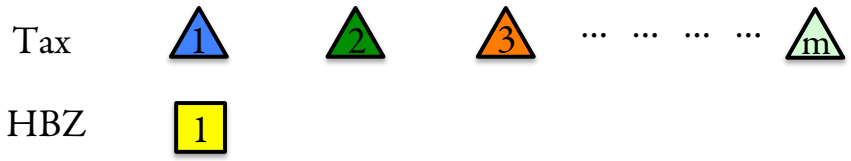
A.4 Proofs and Derivations for the Multi-Locus Model (6.1) in the Special Case When $n = 1$: The '$m \times 1$' Model, Corresponding to One Antigenically Variable Epitope and One Conserved Epitope

Here we take a closer look at the special case $n = 1$ in the multi-locus model (6.1) in which antigenic variation is present at only one of the epitopes. Due to clinical evidence that Tax and HBZ are the key proteins in HTLV-I infection and the experimental observation that there exists a wider degree of variation in Tax (Cavanagh et al., 2006; Gaudray et al., 2002; Kubota et al., 2007; Niewiesk et al., 1994, 1995), we will say that the variable epitope corresponds to the Tax protein while the conserved epitope corresponds to the minus-strand[1] gene product HBZ. However, as with our previous mathematical models, our approach is not specific to Tax and HBZ, but more generally represents any two antigenic targets in which one is variable and the other is conserved. The main purpose of our effort here is to develop several basic ideas that help build a clearer picture of the interactions that occur in the more general multi-locus, or '$m \times n$', model (6.1).

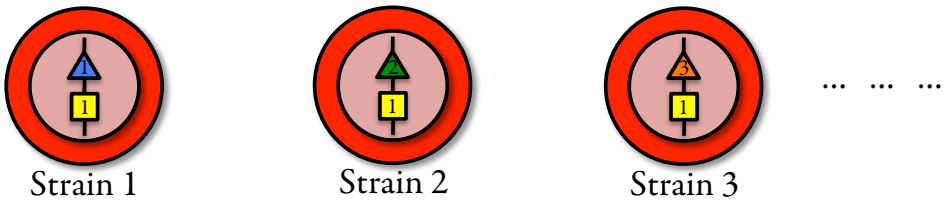
A.4.1 The '$m \times 1$' Model Notation and Equations

Let m be a positive integer and $n = 1$ in the multi-locus model (6.1). Then there are m different variants of the Tax protein, denoted individually by Tax_i, where $i = 1, \ldots, m$. Meanwhile, we assume there is only a single variant of the HBZ protein, denoted simply by HBZ_1. Thus, there are a total of m different strains (see Figure A.1 for an illustration).

[1] Recall that HBZ is the only HTLV-I protein whose RNA is transcribed in the anti-sense direction of the virus genome, i.e. from the $3'$ end to the $5'$ end instead of from $5'$ to $3'$.



(a) Introducing antigenic variation in the Tax epitope.



(b) A virus strain is defined by a combination of each of its epitopes.

Figure A.1: In order to clarify the idea of antigenic variation in the context of our within-host HTLV-I system, we consider an infected cell that expresses two viral epitopes, Tax and HBZ, in which the Tax protein exhibits antigenic variability whilst the HBZ protein is conserved. In the present formulation, because the HBZ protein is common to all proviral cells, each virus strain i is identified only by its Tax variant i.

For simplicity of notation, we will drop the index j. Hence, we consider the following populations:

$x(t)$: healthy (uninfected) $\mathrm{CD4}^+$ helper T-cells,

$u_i(t)$: latently infected $\mathrm{CD4}^+$ helper T-cell with Tax variant i,

$y_i(t)$: actively infected $\mathrm{CD4}^+$ helper T-cell expressing Tax variant i,

$z_i(t)$: $\mathrm{CD8}^+$ CTLs specific to Tax_i,

$s(t)$: $\mathrm{CD8}^+$ CTLs specific to HBZ_1.

The general two-epitope model with antigenic variation in only one epitope, which we refer to as the '$m \times 1$' model, is given by the following $(3m+2)$-dimensional non-linear system of ordinary differential equations (A.15):

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \lambda - \beta x \sum_{j=1}^{m} y_j - \mu_1 x, \\
\frac{\mathrm{d}u_i}{\mathrm{d}t} &= (\beta x + r) y_i - (\tau_i + \mu_2) u_i, \quad \text{for } i = 1, \ldots, m, \\
\frac{\mathrm{d}y_i}{\mathrm{d}t} &= \tau_i u_i - \gamma y_i z_i - \gamma' y_i s - \mu_3 y_i, \quad \text{for } i = 1, \ldots, m, \\
\frac{\mathrm{d}z_i}{\mathrm{d}t} &= \nu \delta_i y_i - \mu_4 z_i, \quad \text{for } i = 1, \ldots, m, \\
\frac{\mathrm{d}s}{\mathrm{d}t} &= \nu' \sum_{j=1}^{m} (1 - \delta_j) y_j - \mu_5 s.
\end{aligned} \tag{A.15}$$

A.4.2 Global Properties of the Infection-Free Equilibrium for the $m \times 1$ Model

The consideration of antigenic variation comes at a cost. Indeed, due to the inherent complexity of multi-strain systems, whereas the model incorporates a greater degree of biological detail, the system also becomes much less tractable mathematically. In particular, we are not able to completely characterise the global dynamical properties of model (A.15) using analytical techniques and can only fully classify the existence and global stability of the infection-free steady state P_0, including defining an invariant region within which the dynamical behaviour can be captured as well as establishing conditions for the instability of P_0. These results, as well as the proofs, are presented in Theorems A.1 and A.2 below.

Theorem A.1 (Biologically Realistic Region for the '$m \times 1$' Model (A.15))**.** *Let* $\mu = \min\{\mu_1, \mu_2, \mu_3\}$*,* $\tilde{\nu} = \max\{\nu, \nu'\}$*, and* $\tilde{\mu} = \min\{\mu_4, \mu_5\}$*, and consider the closed, bounded region*

$$\Gamma_m := \left\{ (x, \vec{u}, \vec{y}, \vec{z}, s) \in \mathbb{R}_+^{3m+2} : x \leq \frac{\lambda}{\mu_1},\ x + \sum_{i=1}^{m}(u_i + y_i) \leq \frac{\lambda}{\mu - r},\ \sum_{i=1}^{m} z_i + s \leq \frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu - r)} \right\}.$$

Assume that $r < \mu$*. Then, the set* Γ_m *defines a positively invariant region with respect to model* (A.15)*. All solutions are bounded for* $t \geq 0$ *and eventually enter* Γ_m*.*

Proof of Theorem A.1. Here we determine a region of interest for model (A.15) for which the dynamical behaviour can be examined. The proof follows a similar argument as for previous models, but is presented here for completeness.

Let $m \geq 1$. Denote $\vec{u} = \{u_i\}_{i=1}^m$, $\vec{y} = \{y_i\}_{i=1}^m$, and $\vec{z} = \{z_i\}_{i=1}^m$. From the equations in model (A.15), it is clear that for any set of non-negative initial conditions, the variables $x(t), \vec{u}(t), \vec{y}(t), \vec{z}(t), s(t)$ remain non-negative for all time $t \geq 0$. As with the previous models, we notice that

$$\frac{\mathrm{d}x}{\mathrm{d}t} \leq \lambda - \mu_1 x,$$

and this implies that

$$\limsup_{t\to\infty} x(t) \leq \frac{\lambda}{\mu_1}.$$

Let $\mu = \min\{\mu_1, \mu_2, \mu_3\}$ and suppose that $r < \mu$. Adding the first $(2m + 1)$ equations of model (A.15) (representing the x, u_i, and y_i compartments for $i = 1, \ldots, m$) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(x + \sum_{i=1}^{m}(u_i + y_i) \right) \leq \lambda - (\mu - r)\left(x + \sum_{i=1}^{m}(u_i + y_i) \right),$$

which implies that $\limsup_{t\to\infty}\left(x+\sum_{i=1}^{m}(u_i+y_i)\right)(t)\leq\lambda/(\mu-r)$. Again, the assumption that $r<\mu$ ensures that the quantity on the right-hand side of the inequality is always positive. To obtain the final inequality, let $\big(x(t),\vec{u}(t),\vec{y}(t),\vec{z}(t),s(t)\big)$ be a solution of model (A.15) with $x(0)+u(0)+y(0)\leq\lambda/(\mu-r)$. Then, summing up all of the CTL populations (both Tax- and HBZ-specific), we compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{m}z_i+s\right)=\sum_{i=1}^{m}\frac{\mathrm{d}z_i}{\mathrm{d}t}+\frac{\mathrm{d}s}{\mathrm{d}t}\leq\tilde{\nu}\sum_{i=1}^{m}y_i-\tilde{\mu}\left(\sum_{i=1}^{m}z_i+s\right),$$

where $\tilde{\nu}=\max\{\nu,\nu'\}$ and $\tilde{\mu}=\min\{\mu_4,\mu_5\}$, so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{m}z_i+s\right)+\tilde{\mu}\left(\sum_{i=1}^{m}z_i+s\right)\leq\frac{\lambda\tilde{\nu}}{\mu-r}\quad\Longrightarrow\quad\limsup_{t\to\infty}\tilde{\mu}\left(\sum_{i=1}^{m}z_i+s\right)\leq\frac{\lambda\nu}{\tilde{\mu}(\mu-r)}.$$

Therefore, we consider the closed, bounded region

$$\Gamma_m:=\left\{(x,\vec{u},\vec{y},\vec{z},s)\in\mathbb{R}_+^{3m+2}:x\leq\frac{\lambda}{\mu_1},\ x+\sum_{i=1}^{m}(u_i+y_i)\leq\frac{\lambda}{\mu-r},\ \sum_{i=1}^{m}z_i+s\leq\frac{\lambda\tilde{\nu}}{\tilde{\mu}(\mu-r)}\right\}.$$

It is a straight-forward exercise to verify that Γ_m is positively invariant in $\mathbb{R}^{3m+2}$ and model (A.15) is well-posed. □

Theorem A.2 (Global Dynamical Properties of the Infection-Free Equilibrium P_0 in the '$m\times 1$' Model)**.** *Let $\vec{0}=(0,\ldots,0)\in\mathbb{R}^m$ denote the m-dimensional zero vector and let $x_H=\lambda/\mu_1$. The infection-free equilibrium for model* (A.15)*, given by $P_0=(x_H,\vec{0},\vec{0},\vec{0},0)$, always exists. If $0<R_{0,i}<1$ for all $i=1,\ldots,m$, then P_0 is the only equilibrium in Γ_m^- and it is globally asymptotically stable in Γ_m^-. Conversely, if there exists $i_0\in\{1,\ldots,m\}$ such that $R_{0,i_0}>1$, then P_0 is unstable. Specifically, let $1\leq k\leq m$ denote the number of viral strains whose basic reproduction number is greater than unity, i.e. $R_{0,i}>1$ for $i=i_1,\ldots,i_k$. Then, P_0 is a saddle with $\dim W^s_{loc}(P_0)=3m-k+2$ and $\dim W^u_{loc}(P_0)=k$, where $W^s_{loc}(P_0)$, $W^u_{loc}(P_0)$ denote the respective local stable and unstable manifolds of P_0.*

Proof. **Part 1. Existence of P_0.**

Let $m\geq 1$. Denote $\vec{u}=\{u_i\}_{i=1}^m$, $\vec{y}=\{y_i\}_{i=1}^m$, and $\vec{z}=\{z_i\}_{i=1}^m$. First, to show the existence of an infection-free steady state P_0, we consider the corresponding equilibrium equations for model (A.15). As with our previous models, in the absence of infection by any strain, that is, if $y_i\equiv 0$ for all $i=1,\ldots,m$, then there exists a steady state representing a healthy individual given by the point $P_0=(x_H,\vec{0},\vec{0},\vec{0},0)\in\mathbb{R}^{3m+2}$, where $\vec{0}=(0,\ldots,0)\in\mathbb{R}^m$ denotes the m-dimensional zero vector and $x_H=\lambda/\mu_1$ is determined by the equilibrium equation corresponding to the x-variable in model (A.15). This infection-free steady state

P_0 always exists.

Part 2. Linear Stability and Instability of P_0.

Next, we demonstrate the local stability properties of P_0, again following a similar argument as with earlier proofs. In particular, we will demonstrate how the basic reproduction numbers, $R_{0,i}$, for each strain $i \in \{1, \ldots, m\}$ establish the linear stability or instability of P_0.

The Jacobian matrix for the general '$m \times 1$' model (A.15) is a $(3m+2)$-dimensional matrix which can be written in block matrix form as

$$J(x, \vec{u}, \vec{y}, \vec{z}, s) = \left[\begin{array}{c|c|c} B_{11} & B_{12} & B_{13} \\ \hline B_{21} & B_{22} & B_{23} \\ \hline B_{31} & B_{32} & B_{33} \end{array}\right].$$

Denote by $\vec{0} \in \mathbb{R}^m$ and $\overline{\mathbf{0}}_{m\times m}$, respectively, the m-dimensional vector and $m \times m$ matrix whose entries are all zero. Also, let $\underset{i=1,\ldots,m}{\text{diag}}\ \{a_i\}$ represent the diagonal matrix with a_i along the main diagonal and zero elsewhere. We can write each of the individual block matrices as follows:

$$B_{11} = \overset{1\times 1 \text{ matrix}}{\left[-\beta \sum_{j=1}^{m} y_j - \mu_1\right]}, \quad B_{12} = \overset{1\times 2m \text{ matrix}}{\left[\ \left[\vec{0}\right]^T \quad \left[\{-\beta x\}_{i=1}^m\right]^T\ \right]},$$

$$B_{13} = \overset{1\times (m+1) \text{ matrix}}{\left[\ \left[\vec{0}\right]^T \qquad 0\ \right]}, \quad B_{21} = \overset{2m\times 1 \text{ matrix}}{\left[\begin{array}{c} \{\beta y_i\}_{i=1}^m \\ \vec{0} \end{array}\right]},$$

$$B_{22} = \overset{2m\times 2m \text{ matrix}}{\left[\begin{array}{cc} \underset{i=1,\ldots,m}{\text{diag}}\ \{-\tau_i - \mu_2\} & \underset{i=1,\ldots,m}{\text{diag}}\ \{\beta x + r\} \\ \underset{i=1,\ldots,m}{\text{diag}}\ \{\tau_i\} & \underset{i=1,\ldots,m}{\text{diag}}\ \{-\gamma z_i - \gamma' s - \mu_3\} \end{array}\right]},$$

$$B_{23} = \overset{2m\times (m+1) \text{ matrix}}{\left[\begin{array}{cc} \overline{\mathbf{0}}_{m\times m} & \vec{0} \\ \underset{i=1,\ldots,m}{\text{diag}}\ \{-\gamma y_i\} & \{-\gamma' y_i\}_{i=1}^m \end{array}\right]}, \quad B_{31} = \overset{(m+1)\times 1 \text{ matrix}}{\left[\begin{array}{c} \vec{0} \\ 0 \end{array}\right]},$$

$$B_{32} = \overset{(m+1)\times 2m \text{ matrix}}{\left[\begin{array}{cc} \overline{\mathbf{0}}_{m\times m} & \underset{i=1,\ldots,m}{\text{diag}}\ \{\nu\delta_i\} \\ \left[\vec{0}\right]^T & \left[\{\nu'(1-\delta_i)\}_{i=1}^m\right]^T \end{array}\right]}, \quad B_{33} = \overset{(m+1)\times (m+1) \text{ matrix}}{\left[\begin{array}{cc} \underset{i=1,\ldots,m}{\text{diag}}\ \{-\mu_4\} & \vec{0} \\ \left[\vec{0}\right]^T & -\mu_5 \end{array}\right]}.$$

Because the block matrices B_{13}, B_{21}, B_{23}, and B_{31} evaluated at the infection-free equilibrium point P_0 are matrices whose entries are all zero, to characterise the linear stability

properties of $J(P_0)$, it suffices to determine the eigenvalues of the diagonal block matrices. We compute the respective characteristic polynomials for B_{11}, B_{22}, and B_{33} as follows:

$$\chi_{B_{11}}(\zeta) = \det\big(\zeta - B_{11}\big) = \zeta + \mu_1,$$

$$\chi_{B_{22}}(\zeta) = \det\big(\zeta I_{2m\times 2m} - B_{22}\big) = \prod_{i=1}^{m}\Big[\zeta^2 + (\tau_i + \mu_2 + \mu_3)\zeta - \mu_3(\tau_i + \mu_2)\big[R_{0,i} - 1\big]\Big],$$

and

$$\chi_{B_{33}}(\zeta) = \det\big(\zeta_{(m+1)\times(m+1)} - B_{33}\big) = (\zeta + \mu_4)^m(\zeta + \mu_5).$$

The roots of the above three characteristic polynomials are precisely the eigenvalues of the original Jacobian matrix at P_0, i.e. $J(P_0)$. Specifically,

$$\zeta_1 = -\mu_1, \quad \text{(1 eigenvalue)},$$

$$\zeta_{2i}, \zeta_{2i+1} = -\frac{1}{2}(\tau_i + \mu_2 + \mu_3) \pm \frac{1}{2}\sqrt{(\tau_i + \mu_2 + \mu_3)^2 + 4\mu_3(\tau_i + \mu_2)[R_{0,i} - 1]},$$
$$\text{for } i = 1, \ldots, m, \quad \text{(2m eigenvalues)},$$

$$\zeta_{2m+i+1} = -\mu_4, \quad \text{for } i = 1, \ldots, m, \quad \text{(m eigenvalues)},$$

and

$$\zeta_{3m+2} = -\mu_5, \quad \text{(1 eigenvalue)}.$$

There are $(2m + 2)$ eigenvalues that are clearly negative, namely, $\zeta_1, \zeta_{3m+2} < 0$ and also $\zeta_{2i+1}, \zeta_{2m+i+1} < 0$ for $i = 1, \ldots, m$. Meanwhile, the remaining m eigenvalues given by ζ_{2i} for $i = 1, \ldots, m$ may be either positive or negative (or zero), depending on the value of the corresponding basic reproduction number for strain i described by $R_{0,i}$. There are two situations that need to be considered. First, suppose that for every $i = 1, \ldots, m$, $R_{0,i} < 1$. Then, $\mathrm{Re}(\zeta_{2i}) < 0$ for each $i = 1, \ldots, m$. Hence, all eigenvalues of $J(P_0)$ are either negative or have negative real part, which means that P_0 is stable and, in particular, P_0 is either a stable node or a stable spiral.

Alternatively, suppose that there exists at least one $i_0 \in \{1, \ldots, m\}$ such that $R_{0,i_0} > 1$. Then it is readily seen that the corresponding eigenvalue, ζ_{2i_0}, has positive real part. In this case, P_0 is unstable. Now, suppose that $1 \leq k \leq m$ of the $R_{0,i}$ are greater than unity. We denote these by the indices $i_i, \ldots, i_k \in \{1, \ldots, m\}$, where $i_i < \ldots < i_k$. The eigenvalues, $\zeta_{2i_1}, \ldots, \zeta_{2i_k}$, corresponding to each of these k strains all have positive real part. Thus, P_0 is a saddle whose local unstable manifold $W^u_{loc}(P_0)$ has dimension k. The local stable manifold $W^s_{loc}(P_0)$ therefore has dimension $(3m - k + 2)$. Lastly, we remark that if $R_{0,j_0} = 1$ for any

$j_0 \in \{1, \ldots, m\}$, then the corresponding eigenvalue ζ_{2j_0} is zero and no immediate conclusions about the local stability of P_0 can be inferred in this case. The preceding argument establishes the linear stability and instability properties of P_0.

Part 3. Global Stability of P_0.

As we have seen above, the infection-free equilibrium P_0 is linearly stable if the basic reproduction number for each strain is less than unity. In fact, such a condition not only implies local stability, but also global stability in the region Γ_m. We will show this result below using the method of Lyapunov functions. The proof again follows a similar argument as in previous models, with slight modifications to suit the current model.

Assume that $R_{0,i} < 1$ for all $i = 1, \ldots, m$. Then by Part 1, P_0 is the only equilibrium point in $\bar{\Gamma}_m$. By Part 2, P_0 is linearly stable. Define a function $L_m : \mathbb{R}_+^{3m+2} \to \mathbb{R}$ as follows:

$$L_m = L_m(x, \vec{u}, \vec{y}, \vec{z}, s) = \left(x - x_H - x_H \log \frac{x}{x_H}\right) + \sum_{i=1}^{m} u_i + \sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\tau_i} y_i + \frac{\gamma}{2\nu} \sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\delta_i \tau_i} z_i^2 + \frac{\gamma'}{2\nu'} \min_{i=1,\ldots,m} \left\{ \frac{\tau_i + \mu_2}{(1-\delta_i)\tau_i} \right\} s^2.$$

The function L_m is positive definite since

(i) $L_m(x_H, \vec{0}, \vec{0}, \vec{0}, 0) = 0$, i.e. $L_m(P_0) = 0$; and

(ii) $L_m(x, \vec{u}, \vec{y}, \vec{z}, s) > 0$, for all $(x, \vec{u}, \vec{y}, \vec{z}, s) \in \mathbb{R}_+^{3m+2}, (x, \vec{u}, \vec{y}, \vec{z}, s) \neq (x_H, \vec{0}, \vec{0}, \vec{0}, 0)$.

The Lyapunov derivative of L_m is

$$\begin{aligned}
\frac{\mathrm{d}L_m}{\mathrm{d}t} &= \left(1 - \frac{x_H}{x}\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \sum_{i=1}^{m} \frac{\mathrm{d}u_i}{\mathrm{d}t} + \sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\tau_i}\frac{\mathrm{d}y_i}{\mathrm{d}t} + \frac{\gamma}{\nu}\sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\delta_i \tau_i} z_i \frac{\mathrm{d}z_i}{\mathrm{d}t} \\
&\quad + \frac{\gamma'}{\nu'} \min_{i=1,\ldots,m} \left\{ \frac{\tau_i + \mu_2}{(1-\delta_i)\tau_i} \right\} s \frac{\mathrm{d}s}{\mathrm{d}t} \\
&= \left(1 - \frac{x_H}{x}\right)(\lambda - \mu_1 x) - \beta x \sum_{i=1}^{m} y_i \left(1 - \frac{x_H}{x}\right) + \sum_{i=1}^{m} \Big[(\beta x + r) y_i - (\tau_i + \mu_2) u_i\Big] \\
&\quad + \sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\tau_i} \Big[\tau_i u_i - \gamma_i y_i z_i - \gamma' y_i s - \mu_3 y_i\Big] + \frac{\gamma}{\nu} \sum_{i=1}^{m} \frac{\tau_i + \mu_2}{\delta_i \tau_i} z_i \Big[\nu \delta_i y_i - \mu_4 z_i\Big] \\
&\quad + \frac{\gamma'}{\nu'} \min_{i=1,\ldots,m} \left\{ \frac{\tau_i + \mu_2}{(1-\delta_i)\tau_i} \right\} s \left[\nu' \sum_{i=1}^{m} (1 - \delta_i) y_i - \mu_5 s\right].
\end{aligned}$$

Replacing $\lambda = \mu_1 x_H$, expanding terms, and noting that $\min_{i=1,\ldots,m} \big\{(\tau_i + \mu_2)/\big((1-\delta_i)\tau_i\big)\big\} \leq (\tau_{i_0} + \mu_2)/\big((1-\delta_{i_0})\tau_{i_0}\big)$ for any fixed $i_0 \in \{1, \ldots, m\}$ (to allow cancellation of several terms)

yields

$$\begin{aligned}\frac{\mathrm{d}L_m}{\mathrm{d}t} &= -\frac{\mu_1}{x}(x-x_H)^2 + \mu_3\sum_{i=1}^{m}\frac{\tau_i+\mu_2}{\tau_i}\Big[R_{0,i}-1\Big]y_i - \frac{\gamma}{\phi_{\mathrm{Tax}}}\sum_{i=1}^{m}\frac{\tau_i+\mu_2}{\delta_i\tau_i}z_i^2 \\ &\quad - \frac{\gamma'}{\phi_{\mathrm{HBZ}}}\min_{i=1,\ldots,m}\left\{\frac{\tau_i+\mu_2}{(1-\delta_i)\tau_i}\right\}s^2 \\ &\leq 0, \quad \forall(x,\vec{u},\vec{y},\vec{z},s)\in\Gamma_m, \quad \text{since } R_{0,i}<1 \quad \text{for all } i=1,\ldots,m.\end{aligned}$$

Hence, L_m is an appropriate Lyapunov function and moreover, $\mathrm{d}L_m/\mathrm{d}t(P_0) = L_m(P_0) = 0$. Since the feasible region Γ_m is compact and invariant with respect to the vector field of model (A.15), it then follows from LaSalle's invariance principle (LaSalle, 1976) that every trajectory starting in Γ_m approaches a set M_m, where M_m is the largest invariant subset of

$$I_m = \left\{(x,\vec{u},\vec{y},\vec{z},s)\in\mathbb{R}_+^{3m+2} : \frac{\mathrm{d}L_m}{\mathrm{d}t}(x,\vec{u},\vec{y},\vec{z},s) = 0\right\},$$

which, in the present case, is precisely

$$\{x\equiv x_H\}\cap\bigcap_{i=1}^{m}\{y_i\equiv 0\}\cap\bigcap_{i=1}^{m}\{z_i\equiv 0\}\cap\{s\equiv 0\}.$$

To establish the global stability of P_0 when $R_0<1$, it remains to show that the set M_m consists solely of the equilibrium point P_0. Observe that in the set I_m, the variable u can take on any arbitrary value along the non-negative real line. Let $\omega_m(t) = \big(x(t),\vec{u}(t),\vec{y}(t),\vec{z}(t),s(t)\big)$ be any trajectory of model (A.15) starting in I_m, so that $x(0)=x_H$, and for $i=1\ldots,m$, $y_i(0)=0$, $z_i(0)=0$, and $s(0)=0$. Note that the value of $u_i(0)=u_{0,i}$ is not yet known at this stage. Then, the behaviour of $\omega_m(t)$ at $t=0$ is governed by

$$\frac{\mathrm{d}x}{\mathrm{d}t}=0, \quad \frac{\mathrm{d}s}{\mathrm{d}t}=0,$$

and for $i=1,\ldots,m$,

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -(\tau_i+\mu_2)u_i, \quad \frac{\mathrm{d}y_i}{\mathrm{d}t}=\tau u_i, \quad \frac{\mathrm{d}z_i}{\mathrm{d}t}=0.$$

Hence, $x(t)\equiv x_H$, $\vec{z}(t)\equiv\vec{0}$, $s(t)\equiv 0$, and

$$u_i(t') = u_i(0)e^{-(\tau_i+\mu_2)t}, \quad \text{for } i=1,\ldots,m.$$

For each $i=1,\ldots,m$, the y_i-compartment varies according to $\mathrm{d}y_i/\mathrm{d}t=\tau u_i(0)$ at time $t=0$. If for any $i_0\in\{1,\ldots,m\}$, $u_{i_0}(0)\neq 0$, then $\mathrm{d}y_{i_0}/\mathrm{d}t>0$, implying that $y_{i_0}(t_0)\neq 0$ for some $t_0\neq 0$; that is, the trajectory will leave the set I_m. For the solution $\omega_m(t)$ to remain in I_m for all t, we must have $u_{i_0}(0)=0$ and therefore $u_{i_0}(t)\equiv 0$. Subsequently, solving the initial value problem $\mathrm{d}y_{i_0}/\mathrm{d}t=0$, $y_{i_0}(0)=0$ yields $y_{i_0}(t)\equiv 0$. As this argument applies to

any choice of i_0, then it holds for every $i \in \{1, \ldots, m\}$ (otherwise, that particular trajectory will leave the set I_m). Taken all together, we obtain that the maximal invariant set in I_m is comprised only of the single point $P_0 = (x_H, \vec{0}, \vec{0}, \vec{0}, 0)$, i.e. $M_m = \{P_0\}$. This completes the proof of Theorem A.2. □

Remark A.3. In the case of chronic infection, we are unable to derive analogous analytical results. In examining the '$m \times 1$' model, we instead choose to break apart the underlying mechanisms of the system by focussing independently on separate aspects of the host-virus interaction, which will allow us to gain several insights analytically. We are able to show the existence of a chronic infection steady state as well as derive several useful relationships among the variables describing the viral and CTL populations by considering two different simplifications, each of which focusses on a certain aspect of the model whilst keeping the other parts of the model fixed. Specifically, in Appendix A.4.3, we first focus on the role of immunogenicity of Tax variants in the absence of competition among viral strains (i.e. $R_{0,i} = R_{0,j}$ for all $i, j \in \{1, \ldots, m\}$). Following this, in Appendix A.4.4, we introduce inter-strain competition (i.e. the $R_{0,i}$'s can be different) and examine its effect on the outcome of chronic infection. Under each of these two simplifications, we are able to prove the existence of at least one chronic infection equilibrium when the basic reproduction number of at least one strain is greater than unity. Furthermore, we have performed an extensive numerical investigation of the model using parameter values within biologically realistic ranges as discussed in Table 3.1. Although not proven, the results of numerical simulations suggest that model (A.15) can only admit one stable steady state at a time, and this has been observed for a wide range of parameter choices.

Remark A.4. In model (A.15), we notice that the equations describing the dynamics for each strain i can be de-coupled from the others. As a result, for each $1 \leq j \leq m$, it is possible for there to exist what we refer to as 'j-strain' equilibria, describing the set of non-negative steady states for which precisely j of the strains are present (i.e. the corresponding components are positive), and the remaining $m-j$ strains are absent (i.e. the corresponding components are zero). Due to the possible combinations for which a fixed number of strains can be present at the same time, for each $j \in \{1, \ldots, m\}$, there are a total of $\binom{m}{j}$ potential j-strain equilibria (note that we are not saying that these equilibria exist and are biologically meaningful; at this stage, we are only considering the possibility according to an initial look at the equations). For example, when $j = 1$, there are a total of $\binom{m}{1} = m$ potential chronic infection equilibrium points representing infection by a single strain. We also highlight that the case $j = m$ corresponds to the scenario in which the chronic infection equilibrium, if it exists, represents the co-existence of all m distinct viral strains simultaneously. As we

branch out into the two different simplifications discussed above in the following sections, we work to show the existence of a chronic infection equilibrium describing co-existence of all possible strains, i.e. the case $j = m$. The reason for this is because the proofs of the existence of any j-strain equilibrium for $1 \leq j \leq m$ follow a similar method of derivation and so it suffices to show the case $j = m$.

A.4.3 Focussing on Immunogenicity

We begin by focussing on elucidating the role of immunogenicity of each Tax variant in the absence of competition. Namely, let us assume that the three factors that govern the success of each strain are the same, so that all virus strains have identical fitness. In other words, we will assume that for every strain, the rates of infectious transmission, mitotic division, and spontaneous infected target cell activation are β, r, and τ, respectively. Then, $R_{0,i} = R_0 = \tau(\beta x + r)/\big(\mu_3(\tau + \mu_2)\big)$ for all $i = 1, \ldots, m$.

A.4.3.1 Coexistence

In the absence of inter-strain competition, all virus strains have identical fitness, and therefore the existence of one strain can be expected to imply the presence of the other. Moreover, the basic reproduction number $R_{0,i}$, which is the same for all strains i, determines whether or not chronic infection is possible.

Formally, we have the following analytical result for the existence of a unique chronic infection state in the situation where all strains are equally fit. The proof follows a similar argument as with our previous models.

Proposition A.5 (Existence of P^* Without Inter-Strain Competition)**.** *Let $R_{0,i} = R_0$ for all $i = 1, \ldots, m$, and suppose that $R_0 > 1$. Then, there exists a unique chronic-infection equilibrium $P^* = (x^*, \vec{u}^*, \vec{y}^*, \vec{z}^*, s^*)$ in $\mathring{\Gamma}_m$, where $\vec{u}^* = (u_1^*, \ldots, u_m^*)$, $\vec{y}^* = (y_1^*, \ldots, y_m^*)$, $\vec{z}^* = (z_1^*, \ldots, z_m^*)$.*

Proof. We will follow a similar method of proof as for the existence of interior equilibria in previous models. Namely, we search for steady state solutions of the form $P^* = (x^*, \vec{u}^*, \vec{y}^*, \vec{z}^*, s^*)$, where $\vec{u}^* = (u_1^*, \ldots, u_m^*)$, $\vec{y}^* = (y_1^*, \ldots, y_m^*)$, $\vec{z}^* = (z_1^*, \ldots, z_m^*)$, and all entries of P^* are non-negative. To simplify the system in this case so as to ignore the effects of inter-strain competition, we assume that all strains have identical rates of spontaneous viral expression, i.e. $\tau_i = \tau$ for all $i = 1, \ldots, m$. Hence, in such a case, $R_{0,i} = R_0$.

Equilibrium points of model (A.15) under the above conditions, if they exist, satisfy the following equations:

$$0 = \lambda - \beta x^* \sum_{j=1}^{m} y_j^* - \mu_1 x^*, \tag{A.16a}$$

$$0 = (\beta x^* + r) y_i^* - (\tau + \mu_2) u_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.16b}$$

$$0 = \tau u_i^* - \gamma y_i^* z_i^* - \gamma' y_i^* s^* - \mu_3 y_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.16c}$$

$$0 = \nu \delta_i y_i^* - \mu_4 z_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.16d}$$

$$0 = \nu' \sum_{j=1}^{m} (1 - \delta_j) y_j^* - \mu_5 s^*. \tag{A.16e}$$

We use these equations to express all of the variables at equilibrium in terms of y_1^*. From Equations (A.16d) and (A.16e), we find that

$$z_i^* = \phi_{\text{Tax}} \delta_i y_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.17}$$

and

$$s^* = \phi_{\text{HBZ}} \sum_{j=1}^{m} (1 - \delta_j) y_j^*, \tag{A.18}$$

where $\phi_{\text{Tax}} = \nu/\mu_4$ and $\phi_{\text{HBZ}} = \nu'/\mu_5$.

Next, substituting the above expressions into Equation (A.16c) yields

$$u_i^* = \frac{1}{\tau} \left(\gamma \phi_{\text{Tax}} \delta_i y_i^* + \gamma' \phi_{\text{HBZ}} \sum_{j=1}^{m} (1 - \delta_j) y_j^* + \mu_3 \right) y_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.19}$$

Define the function

$$R_(x) = \frac{\tau}{\mu_3 (\tau + \mu_2)} \big(\beta x + r \big). \tag{A.20}$$

Observe that at the level of healthy cells, given by x_H, at the infection-free equilibrium P_0, we obtain $R(x_H) = R_0$, which is identical to the expression for the basic reproduction number in previous models (c.f. the expressions that appear in Equations (3.2), (4.2), and (5.2)).

After substitution of Equations (A.17)–(A.19) into Equation (A.16b) and re-arranging terms, we derive the following relationship:

$$\mu_3 \big[R(x^*) - 1 \big] = \gamma \phi_{\text{Tax}} \delta_i y_i^* + \gamma' \phi_{\text{HBZ}} \sum_{j=1}^{m} (1 - \delta_j) y_j^*, \quad \text{for each } i = 1, \ldots, m. \tag{A.21}$$

Notice that the term on the left side as well as the second term on the right side of Equation (A.21) are equal for any choice of $i \in \{1, \dots, m\}$. Thus, we can conclude that

$$\delta_{i_1} y^*_{i_1} = \delta_{i_2} y^*_{i_2}, \quad \text{for any } i_1, i_2 \in \{1, \dots, m\}. \tag{A.22}$$

In other words, every strain can be related to each other. Next, from Equation (A.16a), we have the following relationship between x^* and the sum of all strains y^*_j:

$$\frac{\mu_1}{\beta}\left(\frac{x_H}{x^*} - 1\right) = \sum_{j=1}^{m} y^*_j. \tag{A.23}$$

Note that since we are considering non-negative values for all y^*_j, where $j \in \{1, \dots, m\}$, then according to Equation (A.23), it follows that $x^* \leq x_H$.

Next, substitution of Equation (A.22) into Equation (A.23) yields the following relation, which is now only in terms of x^* and y^*_i, the latter variable representing any single particular strain $i \in \{1, \dots, m\}$ (rather than the sum of all possible strains):

$$\delta_i y^*_i = \frac{\mu_1}{\beta}\left(\sum_{j=1}^{m} \frac{1}{\delta_j}\right)^{-1}\left(\frac{x_H}{x^*} - 1\right). \tag{A.24}$$

By substituting Equations (A.22) and (A.24) for each of the strains in Equation (A.21), we obtain for each $i = 1, \dots, m$:

$$\mu_3\big[R(x^*) - 1\big] = \frac{\mu_1}{\beta}\frac{\left[\gamma\phi_{\text{Tax}} + \gamma'\phi_{\text{HBZ}}\sum_{j=1}^{m}\frac{(1-\delta_j)}{\delta_j}\right]}{\sum_{j=1}^{m}\frac{1}{\delta_j}}\left(\frac{x_H}{x^*} - 1\right). \tag{A.25}$$

Thus, a chronic infection equilibrium P^* would satisfy the above Equation (A.25). Proving that such a point exists is equivalent to showing the existence of a positive root of the following function:

$$F(x) = \big[R(x) - 1\big] + \frac{\mu_1}{\beta\mu_3}\frac{\left[\gamma\phi_{\text{Tax}} + \gamma'\phi_{\text{HBZ}}\displaystyle\sum_{j=1}^{m}\frac{(1-\delta_j)}{\delta_j}\right]}{\displaystyle\sum_{j=1}^{m}\frac{1}{\delta_j}}\left(1 - \frac{x_H}{x}\right). \tag{A.26}$$

We compute

$$F'(x) = R'(x) + \frac{\mu_1}{\beta\mu_3}\frac{\left[\gamma\phi_{\text{Tax}} + \gamma'\phi_{\text{HBZ}}\sum_{j=1}^{m}\frac{(1-\delta_j)}{\delta_j}\right]}{\sum_{j=1}^{m}\frac{1}{\delta_j}}\frac{x_H}{x^2},$$

where $R'(x) = \beta\tau/\big(\mu_3(\tau + \mu_2)\big) > 0$. Hence, $F'(x) > 0\ \forall x \neq 0$, i.e. $F(x)$ is a monotonically increasing function of x. Moreover, it is straight-forward to compute

$$\lim_{x\to 0^+} F(x) = -\infty \quad \text{and} \quad F(x_H) = \big[R_0 - 1\big].$$

Then, the existence of a unique positive root, x^*, of the equation $F(x) = 0$ is equivalent to the condition that $R_0 > 1$.

Therefore, in the case where there is no competition among strains, the chronic infection steady state P^* exists if and only if $R_0 > 1$, where R_0 represents the basic reproduction number for all strains. This concludes the proof. □

We have not been able to establish analytically the stability properties of the chronic infection equilibrium using the methodology developed for previous models. However, we have performed extensive computational investigation of this problem, namely, examining the behaviour of the system as the value of R_0 passes unity (i.e. by perturbing the parameters found in the expression for R_0), including numerically determining the eigenvalues of the Jacobian matrix at the chronic infection steady state P^*. The results of these simulations, which have not been shown, are consistent with the stability properties of the chronic infection equilibrium in our previous models, specifically that, for a wide range of parameter choices, all eigenvalues appear to have negative real part whenever $R_0 > 1$, and so this condition implies that the healthy state P_0 is unstable and the chronic infection state P^* is stable.

A.4.3.2 The extent of Tax immunogenicity determines the relative abundances of virus strains and CTLs

Despite the various virus strains having equal fitness, a detailed analytical exploration of the stable co-existence equilibrium has demonstrated that the relative abundances of individual virus strains and epitope-specific CTLs are heavily influenced by the degree of immunogenicity towards each of the m Tax variants. Furthermore, these quantities are independent of factors arising due to host immunity and viral fitness.

As long as all of the δ_i, where $i = 1, \ldots, m$, are non-zero[2], we can derive the following relationships for the viral and CTL populations at equilibrium. A detailed derivation of each of the expressions follows. Specifically, for all $i_1, i_2 \in \{1, \ldots, m\}$, we have for the virus population:

(V1) Ratio of active virus strain i_1 versus strain i_2: $\delta_{i_1} y^*_{i_1} = \delta_{i_2} y^*_{i_2}$;

(V2) Ratio of latent virus strain i_1 versus strain i_2: $\frac{u^*_{i_1}}{u^*_{i_2}} = \frac{y^*_{i_1}}{y^*_{i_2}}$, i.e. $\delta_{i_1} u^*_{i_1} = \delta_{i_2} u^*_{i_2}$;

(V3) Total virus population: $\sum_{j=1}^{m} (u^*_j + y^*_j) = \delta_{i_1} \Big(\sum_{j=1}^{m} \frac{1}{\delta_j} \Big) (u^*_{i_1} + y^*_{i_1})$;

[2] If $\delta_{i_0} = 0$ for a given $i_0 \in \{1, \ldots, m\}$, then the corresponding equation for the CTL type targeting Tax variant i_0 would represent exponential decay, meaning that this CTL population would disappear at equilibrium and hence the dimension of the system would be reduced.

(V4) Therefore, the relative abundances of each viral strain in terms of the total virus population are:

$$\frac{\text{Strain i}}{\text{Total Virus}} = \frac{u_{i_1}^* + y_{i_1}^*}{\sum_{j=1}^m (u_j^* + y_j^*)} = \frac{1/\delta_{i_1}}{\sum_{j=1}^m \frac{1}{\delta_j}}.$$

and for the CTL population:

(C1) Tax variant-specific CTLs are in equal abundance: $z_{i_1}^* = z_{i_2}^*$;

(C2) Total CTL population: $\left(\sum_{j=1}^m z_j^*\right) + s^* = \left[m\phi_{\text{Tax}} + \phi_{\text{HBZ}} \sum_{j=1}^m \left(\frac{1-\delta_j}{\delta_j}\right)\right]\delta_1 y_1^*$;

(C3) The dominance of a CTL type in terms of the ratio of Tax-specific CTLs to HBZ-specific CTLs is given by:

$$\frac{\text{CTL}_{\text{Tax}}}{\text{CTL}_{\text{HBZ}}} = \frac{\sum_{j=1}^m z_j^*}{s^*} = \frac{m\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} \left(\sum_{j=1}^m \frac{1-\delta_j}{\delta_j}\right)^{-1}. \tag{A.27}$$

From Equation (A.27), we observe that the dominance of CTLs targeting Tax or HBZ is determined only by parameters defining the number of distinct Tax variants, their respective immunogenicities, and the ratio of effective expansion rates of Tax- versus HBZ-specific CTL responses. This observation in turn allows us to establish whether the bulk of the CTL response is strain-specific and targeting the Tax epitope, or whether it is targeting instead the shared HBZ eptiope. In particular, there are three mutually exclusive outcomes for the CTL response:

(a) $\sum_{j=1}^m \frac{1}{\delta_j} < m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} + 1\right)$: Tax-specific CTLs eventually dominate;

(b) $\sum_{j=1}^m \frac{1}{\delta_j} = m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} + 1\right)$: The total number of Tax-specific CTLs (consisting of all variant-specific CTL types) exists at equal abundance as the compartment of HBZ-specific CTLs at equilibrium;

(c) $\sum_{j=1}^m \frac{1}{\delta_j} > m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} + 1\right)$: HBZ-specific CTLs eventually dominate.

Proof. In the following, we derive the various expressions relating viral and CTL populations that we have stated above, i.e. properties **(V1)**–**(V4)** for the former and properties **(C1)**–**(C3)** for the latter. The system under consideration is the simplified version of model (A.15) in which all strains have identical fitness, i.e. $\tau_i = \tau$ for all $i = 1, \ldots, m$ (note that this implies that $R_{0,i} = R_0$ for all $i = 1, \ldots, m$). Assume that a chronic infection equilibrium $P^* =$

$(x^*, \vec{u}^*, \vec{y}^*, \vec{z}^*, s^*)$ for this system exists, where $\vec{u}^* = \{u_i^*\}_{i=1}^m$, $\vec{y}^* = \{y_i^*\}_{i=1}^m$, $\vec{z}^* = \{z_i^*\}_{i=1}^m$, and all entries in P^* are non-negative.

Let $i_1, i_2 \in \{1, \ldots, m\}$ represent the index of any two virus strains. For the virus population:

(V1) Equation (A.22) shows the relationship between strain i_1 and strain i_2. Namely,

$$\delta_{i_1} y_{i_1}^* = \delta_{i_2} y_{i_2}^*. \tag{A.28}$$

(V2) From Equation (A.19), the latent compartment for strains i_1 and i_2 are, respectively,

$$u_{i_1}^* = \frac{1}{\tau}\left(\gamma\phi_{\text{Tax}}\delta_{i_1} y_{i_1}^* + \gamma'\phi_{\text{HBZ}}\sum_{j=1}^m (1-\delta_j) y_j^* + \mu_3\right) y_{i_1}^*,$$

and

$$u_{i_2}^* = \frac{1}{\tau}\left(\gamma\phi_{\text{Tax}}\delta_{i_2} y_{i_2}^* + \gamma'\phi_{\text{HBZ}}\sum_{j=1}^m (1-\delta_j) y_j^* + \mu_3\right) y_{i_2}^*,$$

Substituting Equation (A.28) into the first term of either equation, we observe that the expression in parentheses is equal for both $u_{i_1}^*$ and $u_{i_2}^*$, hence we can determine the ratio of latent virus strain i_1 versus strain i_2 to be

$$\frac{u_{i_1}^*}{u_{i_2}^*} = \frac{y_{i_1}^*}{y_{i_2}^*}.$$

Thus, again by Equation (A.28),

$$\delta_{i_1} u_{i_1}^* = \delta_{i_2} u_{i_2}^*. \tag{A.29}$$

(V3) The total virus population is computed by applying Equations (A.28) and (A.29) to the definition of total virus as the sum of all provirus-positive cells, both latent and active, across the possible strains. That is, Total Virus $= \sum_{j=1}^m (u_j^* + y_j^*)$. Thus, we have

$$\text{Total Virus} = \delta_{i_1}\left(\sum_{j=1}^m \frac{1}{\delta_j}\right)\left(u_{i_1}^* + y_{i_1}^*\right). \tag{A.30}$$

Note, in particular, that Equation (A.30) is now only dependent on a single virus strain.

(V4) Using Equation (A.30), we can calculate the relative abundance of strain $i_1 \in \{1, \ldots, m\}$ within the virus-infected cell pool as

$$\frac{\text{Strain } i_1}{\text{Total Virus}} = \frac{u_{i_1}^* + y_{i_1}^*}{\delta_{i_1}\left(\sum_{j=1}^m \frac{1}{\delta_j}\right)\left(u_{i_1}^* + y_{i_1}^*\right)} = \frac{1/\delta_{i_1}}{\sum_{j=1}^m \frac{1}{\delta_j}}.$$

Next, for the CTL population:

(C1) Substitution of Equation (A.28) into Equation (A.17) yields

$$z_{i_1}^* = z_{i_2}^* \quad \text{for any } i_1, i_2 \in \{1, \ldots, m\}. \tag{A.31}$$

(C2) The total CTL population is obtained by summing the z_j^*, for $j = 1, \ldots, m$, using Equation (A.17) and s^* from Equation (A.18); that is, Total CTL $= \sum_{j=1}^{m} z_j^* + s^*$. Then, substituting Equation (A.28) appropriately, we obtain

$$\begin{aligned}\text{Total CTL} &= \phi_{\text{Tax}} \sum_{j=1}^{m} \delta_j y_j^* + \phi_{\text{HBZ}} \sum_{j=1}^{m} (1-\delta_j) y_j^* \\ &= \phi_{\text{Tax}} \delta_{i_1} y_{i_1}^* \sum_{j=1}^{m} 1 + \phi_{\text{HBZ}} \delta_{i_1} y_{i_1}^* \sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j}.\end{aligned}$$

Note that in the above equations, the top line is a sum of all possible strains, whereas the subsequent line is only dependent on the single strain i_1. This leaves us with the relation

$$\text{Total CTL} = \left[m\phi_{\text{Tax}} + \phi_{\text{HBZ}} \sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j} \right] \delta_{i_1} y_{i_1}^*. \tag{A.32}$$

(C3) We compute the abundances of Tax-specific CTLs (all types) and HBZ-specific CTLs using the respective terms in Equation (A.32) as follows:

$$\frac{\text{CTL}_{\text{Tax}}}{\text{CTL}_{\text{Total}}} = \frac{m\phi_{\text{Tax}}}{m\phi_{\text{Tax}} + \phi_{\text{HBZ}} \sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j}} \tag{A.33}$$

and

$$\frac{\text{CTL}_{\text{HBZ}}}{\text{CTL}_{\text{Total}}} = \frac{\phi_{\text{HBZ}} \sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j}}{m\phi_{\text{Tax}} + \phi_{\text{HBZ}} \sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j}}. \tag{A.34}$$

Moreover, the two epitope-specific CTL responses (i.e. Tax versus HBZ, where all Tax variant-specific CTL responses have been grouped together) are in equal abundance when

$$\frac{\text{CTL}_{\text{Tax}}}{\text{CTL}_{\text{HBZ}}} = \frac{m\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} \left(\sum_{j=1}^{m} \frac{(1-\delta_j)}{\delta_j} \right)^{-1} = 1$$

if and only if

$$\sum_{j=1}^{m}\left(\frac{1}{\delta_j}-1\right)=\frac{m\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}$$

if and only if

$$\sum_{j=1}^{m}\frac{1}{\delta_j}=m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}+1\right). \tag{A.35}$$

Similarly, using Equations (A.33) and (A.34), we can solve the inequalities when the ratio of Tax-specific to HBZ-specific CTLs is greater than unity (in which case, Tax-specific CTLs are in greater abundance than HBZ-specific) or, conversely, less than unity. All together, there are three possibilities that characterise the dominance of each epitope-specific CTL response. In particular,

(a) If $\displaystyle\sum_{j=1}^{m}\frac{1}{\delta_j}<m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}+1\right)$, then Tax-specific CTLs dominate;

(b) If $\displaystyle\sum_{j=1}^{m}\frac{1}{\delta_j}=m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}+1\right)$, then both Tax-specific CTLs (consisting of all types) and HBZ-specific CTLs exist at equal abundances;

(c) If $\displaystyle\sum_{j=1}^{m}\frac{1}{\delta_j}>m\left(\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}}+1\right)$, then HBZ-specific CTLs dominate.

This completes the derivation of all the properties as stated.

□

Remark A.5 (A Special Case)**.** We consider the following special case for model (A.15) without the effects of inter-strain competition, namely, when δ_i are equal for all $i=1,\ldots,m$, which we denote simply by δ.

Then, by **(V1)**, all virus strains are at equal abundance. The relative abundances of CTLs targeting either one of the two epitopes, Tax or HBZ, depends on the value of the single parameter δ as in the basic two-epitope model (4.1). Specifically, the dominance of Tax-specific and HBZ-specific CTLs for this special case are, respectively, given by the following expressions (see the derivation of **(C3)**, namely, Equations (A.33) and (A.34); also, compare with Equation (4.5)):

$$\frac{\text{CTL}_{\text{Tax}}}{\text{CTL}_{\text{Total}}}=\frac{\delta\phi_{\text{Tax}}}{\delta\phi_{\text{Tax}}+(1-\delta)\phi_{\text{HBZ}}}\quad\text{and}\quad\frac{\text{CTL}_{\text{HBZ}}}{\text{CTL}_{\text{Total}}}=\frac{(1-\delta)\phi_{\text{HBZ}}}{\delta\phi_{\text{Tax}}+(1-\delta)\phi_{\text{HBZ}}}.$$

At the endpoints for δ, we briefly note that the above expressions imply the following:

(a) If $\delta = 1$, then anti-HTLV-I host immunity consists only of CTLs targeting specific Tax variants. The immune response is therefore completely strain-specific;

(b) If $\delta = 0$, then host immunity is comprised only of CTLs targeting the shared HBZ epitope. The immune response is therefore fully cross-reactive.

A.4.4 Focussing on Competition Between Strains

The basic reproduction number $R_{0,i}$ defines viral fitness for Strain i and distinguishes the different strains. We have assumed that a virus strain is able to gain a competitive advantage over other strains if latently infected cells spontaneously become re-activated at an increased rate. From the proof of Theorem A.2, we can deduce that a sufficient condition for the extinction of any given virus strain is $R_{0,i} < 1$. Without competitive effects between strains, there is a single R_0 value representing the fitness of all virus strains, and such a condition is also necessary for convergence to the infection-free steady state. As we are interested here in the case where virus strains differ in fitness, to examine the effects of inter-strain competition, we will assume that all virus strains i, for $i = 1, \ldots, m$, have values of $R_{0,i}$ larger than 1.

In the previous section, we have seen that in the absence of competition between strains, the chronic infection state to which the system naturally tends is one where there is stable co-existence of both strains. Moreover, the relative abundances of both the virus and CTL populations are determined by the extent of immunogenicity towards the particular Tax variant possessed by a given virus strain. However, what would happen in the more general case if one of the virus strains has a competitive advantage over the other?

In order to highlight the fundamental principles of inter-strain competition, we let $m = 2$ in model (A.15) and focus our attention on the case where two variants of Tax exist (Tax_1 and Tax_2) alongside a single form of HBZ. The resulting '2×1' model is the simplest way in which we can incorporate antigenic variation in one of the epitopes. The concepts illustrated by this simpler system can be applied to the more general '$m \times 1$' system with any number of Tax variants. With the introduction of inter-strain competition, two possible stable chronic infection outcomes are now possible: (i) co-existence of both strains, and (ii) dominance of the advantageous strain accompanied by the exclusion of the weaker strain. The outcome of such an interaction between strains depends on the extent to which the advantageous strain stimulates cross-reactive host immune responses.

A.4.4.1 The extent of immunogenicity of the advantageous strain determines co-existence or competitive exclusion of virus strains

Without loss of generality, let us assume that strain 1 has a competitive advantage over strain 2. From the expression for the basic reproduction number, such a fitness advantage could be manifested as strain 1 having a greater potential for infectious transmissibility, a faster rate of mitotic spread, or a higher rate of spontaneous activation from the latently infected state. In other words, we assume that $R_{0,1} > R_{0,2} > 1$. In the following, we derive a necessary, though not sufficient, condition for the co-existence of both strains.

Proposition A.6 (Existence of P^* in the Presence of Inter-Strain Competition)**.** *Let $R_{0,1} > R_{0,2} > 1$ and suppose that a chronic infection equilibrium P^* exists. Then, there exists a (unique and fixed) degree of immunogenicity $0 < \tilde{\delta}_1 < 1$ towards Tax_1 (corresponding to Strain 1, the advantageous strain) such that $\delta_1 > \tilde{\delta}_1$ implies the co-existence of both viral strains. Conversely, when $\delta_1 < \tilde{\delta}_1$, the co-existence equilibrium does not exist.*

Proof. Let $R_{0,1} > R_{0,2} > 1$ and suppose that there exists a unique chronic infection steady state $P^* = (x^*, u_1^*, u_2^*, y_1^*, y_2^*, z_1^*, z_2^*, s^*)$ whose compartments satisfy the following equilibrium equations for model (A.15) with $m = 2$ strains:

$$0 = \lambda - \beta x^* \sum_{j=1}^{m} y_j^* - \mu_1 x^*, \tag{A.36a}$$

$$0 = (\beta x^* + r) y_i^* - (\tau_i + \mu_2) u_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.36b}$$

$$0 = \tau_i u_i^* - \gamma y_i^* z_i^* - \gamma' y_i^* s^* - \mu_3 y_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.36c}$$

$$0 = \nu \delta_i y_i^* - \mu_4 z_i^*, \quad \text{for } i = 1, \ldots, m, \tag{A.36d}$$

$$0 = \nu' \sum_{j=1}^{m} (1 - \delta_j) y_j^* - \mu_5 s^*. \tag{A.36e}$$

By Equations (A.36b), (A.36d) and (A.36e),

$$u_i^* = \frac{1}{\tau_i + \mu_2} (\beta x + r) y_i^* \quad \text{and} \quad z_i = \phi_{\text{Tax}} \delta_i y_i^*, \quad \text{for } i = 1, 2, \tag{A.37}$$

and

$$s^* = \phi_{\text{HBZ}} \big[(1 - \delta_1) y_1^* + (1 - \delta_2) y_2^* \big]. \tag{A.38}$$

By Equation (A.36a),

$$y_1^* + y_2^* = \frac{\mu_1}{\beta} \left(\frac{x_H}{x^*} - 1 \right). \tag{A.39}$$

As with previous models, we require $0 < x^* < x_H$, otherwise from Equation (A.39), $y_1^* + y_2^* < 0$, which is biologically impossible. Also, if $x^* = x_H$, then $y_1^* + y_2^* = 0$ and no infection occurs.

Substitution of Equations (A.37)–(A.38) into Equation (A.36c) yields

$$\gamma\phi_{\text{Tax}}\delta_i y_i^* + \gamma'\phi_{\text{HBZ}}\big[(1-\delta_1)y_1^* + (1-\delta_2)y_2^*\big] = \mu_3\big[R_i(x^*) - 1\big], \quad \text{for } i = 1, 2. \tag{A.40}$$

Note that Equation (A.40) implies that $R_i(x^*) > 1$ for $i = 1, 2$ (i.e. because the left-hand side of the equation is positive, so must be the right-hand side). We can write Equation (A.40) as $\mathbf{Ay} = \mathbf{b}$:

$$\underbrace{\begin{bmatrix} \gamma\phi_{\text{Tax}}\delta_1 + \gamma'\phi_{\text{HBZ}}(1-\delta_1) & \gamma'\phi_{\text{HBZ}}(1-\delta_2) \\ \gamma'\phi_{\text{HBZ}}(1-\delta_1) & \gamma\phi_{\text{Tax}}\delta_2 + \gamma'\phi_{\text{HBZ}}(1-\delta_2) \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} y_1^* \\ y_2^* \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \mu_3\big[R_1(x) - 1\big] \\ \mu_3\big[R_2(x) - 1\big] \end{bmatrix}}_{\mathbf{b}}. \tag{A.41}$$

The determinant of the coefficient matrix $\mathbf{A}$ can be computed as

$$\det \mathbf{A} = \big(\gamma\phi_{\text{Tax}}\big)^2\delta_1\delta_2 + \big(\gamma\phi_{\text{Tax}}\big)\big(\gamma'\phi_{\text{HBZ}}\big)\big(\delta_1(1-\delta_2) + \delta_2(1-\delta_1)\big). \tag{A.42}$$

We notice that $\det \mathbf{A} > 0$ since $0 < \delta_1, \delta_2 < 1$. Hence, the matrix $\mathbf{A}$ is invertible and its inverse $\mathbf{A^{-1}}$ is given by

$$\mathbf{A^{-1}} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} \gamma\phi_{\text{Tax}}\delta_2 + \gamma'\phi_{\text{HBZ}}(1-\delta_2) & -\gamma'\phi_{\text{HBZ}}(1-\delta_2) \\ -\gamma'\phi_{\text{HBZ}}(1-\delta_1) & \gamma\phi_{\text{Tax}}\delta_1 + \gamma'\phi_{\text{HBZ}}(1-\delta_1) \end{bmatrix}. \tag{A.43}$$

Hence, the solution of system (A.41) is uniquely determined by $\mathbf{y}^* = \mathbf{A^{-1}b}$, whose components are given by

$$y_1^* = \frac{\mu_3}{\det \mathbf{A}}\Big[\gamma\phi_{\text{Tax}}\delta_2\big[R_1(x^*) - 1\big] + \gamma'\phi_{\text{HBZ}}(1-\delta_2)\big[R_1(x^*) - R_2(x^*)\big]\Big], \tag{A.44}$$

and

$$y_2^* = \frac{\mu_3}{\det \mathbf{A}}\Big[\gamma\phi_{\text{Tax}}\delta_1\big[R_2(x^*) - 1\big] + \gamma'\phi_{\text{HBZ}}(1-\delta_1)\big[R_2(x^*) - R_1(x^*)\big]\Big]. \tag{A.45}$$

The signs of y_1^* and y_2^* depend on the sum within the large, outer square brackets in Equations (A.44) and (A.45). Recall that we have assumed that $R_{0,1} > R_{0,2} > 1$, which is equivalent to $\tau_1 > \tau_2$, and so the following also holds:

$$R_1(x) > R_2(x) \quad \forall x > 0.$$

Thus,

$$R_1(x^*) - R_2(x^*) > 0 \quad \text{but} \quad R_2(x) - R_1(x^*) < 0.$$

We now derive a condition for the co-existence of the two strains. For the advantageous strain (i.e. strain 1), we observe from Equation (A.44) that

$$y_1^* > 0 \quad \text{for any choice of } 0 < \delta_1, \delta_2 < 1.$$

However, this is not true for y_2^*. Instead, using Equation (A.45), we calculate that at $\delta_1 = 0$,

$$y_2^*\big|_{\delta_1=0} = \frac{\mu_3}{\det \mathbf{A}} \gamma' \phi_{\text{HBZ}} \big[R_2(x^*) - R_1(x^*)\big] < 0,$$

while at $\delta_1 = 1$,

$$y_2^*\Big|_{\delta_1=1} = \frac{\mu_3}{\det \mathbf{A}} \gamma \phi_{\text{Tax}} \big[R_2(x^*) - 1\big] > 0,$$

since $R_2(x^*) > 1$. Then, as y_2^* is a smooth function with respect to the parameter δ_1, by the Intermediate Value Theorem,

$$\exists\ \tilde{\delta}_1 \in (0,1) \quad \text{such that} \quad y_2^*\big|_{\delta_1=\tilde{\delta}_1} = 0.$$

That is, $\tilde{\delta}_1$ is a threshold for the extent of immunogenicity of the Tax variant possessed by the advantageous strain 1 below which competitive exclusion occurs. Moreover, this value $\tilde{\delta}_1$ is unique since y_2^* is linear in δ_1. There is an intuitive immunological explanation for the existence of such a threshold parameter. In particular, on the one hand, when $\delta_1 > \tilde{\delta}_1$, immunity is primarily strain-specific with little cross-interference with the lesser fit strain 2. On the other hand, when $\delta_1 < \tilde{\delta}_1$, the advantageous strain 1 stimulates a vigorous cross-reactive CTL response targeting the shared HBZ epitope, and this shared immune response is sufficiently strong to eliminate the lesser fit strain 2.

Lastly, re-arranging Equation (A.45) allows us to determine an explicit expression for $\tilde{\delta}_1$, namely, by requiring that $y_2^* > 0$ as follows:

$$\begin{aligned}
&\gamma \phi_{\text{Tax}} \delta_1 \big[R_2(x^*) - 1\big] + \gamma' \phi_{\text{HBZ}} (1-\delta_1) \big[R_2(x^*) - R_1(x^*)\big] > 0 \\
\text{iff} \quad &\delta_1 \Big[\gamma \phi_{\text{Tax}} \big[R_2(x^*) - 1\big] + \gamma' \phi_{\text{HBZ}} \big[R_1(x^*) - R_2(x^*)\big]\Big] > \gamma' \phi_{\text{HBZ}} \big[\underbrace{R_1(x^*) - R_2(x^*)}_{>0}\big] \\
\text{iff} \quad &\delta_1 > \frac{\gamma' \phi_{\text{HBZ}} \big[R_1(x^*) - R_2(x^*)\big]}{\gamma \phi_{\text{Tax}} \big[R_2(x^*) - 1\big] + \gamma' \phi_{\text{HBZ}} \big[R_1(x^*) - R_2(x^*)\big]}.
\end{aligned}$$

Set

$$\tilde{\delta}_1 := \frac{\gamma' \phi_{\text{HBZ}} \big[R_1(x^*) - R_2(x^*)\big]}{\gamma \phi_{\text{Tax}} \big[R_2(x^*) - 1\big] + \gamma' \phi_{\text{HBZ}} \big[R_1(x^*) - R_2(x^*)\big]}. \tag{A.46}$$

We note that assuming that $R_1(x^*), R_2(x^*) > 1$ (which we have done here), if $R_1(x^*) > R_2(x^*)$, then $0 < \tilde{\delta}_1 < 1$. Hence, in the presence of competition, such a threshold parameter $\tilde{\delta}_1$ always exists, i.e. competitive exclusion is always possible. This concludes the proof. □

In the proof of Proposition A.6, we show that $\tilde{\delta}_1$ is a unique, fixed quantity that depends on the parameters, and hence on the specific values of $R_{0,i}$ for $i = 1, 2$, as well as the level of healthy target cells at the chronic infection co-existence equilibrium. Unfortunately, we have not been able to find satisfactorily intuitive sufficient conditions for the co-existence of both viral strains, nor can we prove the stability of the chronic infection steady state using the analytical methods developed earlier. However, extensive numerical simulations suggest that, like the model simplification considered in the previous section (i.e. in the absence of inter-strain competitive effects), the establishment of persistent infection by at least one strain only requires that one of the $R_{0,i}$ (here $i = 1$ or 2, but this can be extended to any $m > 1$) is greater than unity (results not shown). Furthermore, our numerical results appear to indicate that the quantity $\tilde{\delta}_1$ as stated in Proposition A.6 acts as a threshold parameter determining not only the co-existence of both strains, but also stability — in other words, whenever such a co-existence equilibrium exists, it is stable, at least locally. Taken together, our simulation results also imply that if $\delta_1 < \tilde{\delta}_1$, then the weaker strain 2 is driven to extinction by the principle of competitive exclusion (Hardin, 1960) (results not shown).

More generally, this demonstrates that in the presence of inter-strain fitness differences, immunogenicity towards a common or shared epitope can lead to the extinction of strains that are under strong selection by a broad neutralising host immunity, but do not have the ability to expand rapidly enough to avoid elimination. As we see in Chapter 6, when we incorporate antigenic variation at both epitopes, the introduction of heterogeneity in the fitnesses of even just one strain can lead to a self-organisation of the virus population into ordered sets in which viral strains form a stable structure consisting of a proper subset of all possible virus variants.

A.5 Proofs for the General Multi-Locus Model (6.1)

A.5.1 Derivation of the Expression Relating Equal CTL Frequencies from Section 6.3

Here we derive the mathematical expression describing the relationship between the immunogenicities δ_i, where $i = 1, \ldots, m$, of the various Tax variants, in which both types of epitope-specific CTLs, Tax-specific and HBZ-specific, are functionally identical and achieve equilibrium at equal abundances.

For $k = 1, \ldots, m$ and $l = 1, \ldots, n$, the fitness of each strain kl is determined by its basic reproduction number as given by Equation (6.2):

$$R_{0,kl} = \frac{\tau_{kl}}{\mu_3(\tau_{kl} + \mu_2)}\big(\beta x_H + r\big), \quad \text{where } x_H = \frac{\lambda}{\mu_1}.$$

A chronic infection equilibrium $P^* = (x^*, u_{kl}^*, y_{kl}^*, z_k^*, s_l^*)$ for model (6.1) with $k = 1, \ldots, m$ and $l = 1, \ldots, n$, if it exists, must satisfy the steady state equations

$$0 = \lambda - \beta x^* \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}^* - \mu_1 x^*, \tag{A.47a}$$

$$0 = (\beta x^* + r) y_{kl}^* - (\tau_{kl} + \mu_2) u_{kl}^*, \quad \text{for } k = 1, \ldots, m, \text{ and } l = 1, \ldots, n, \tag{A.47b}$$

$$0 = \tau_{kl} u_{kl}^* - \gamma y_{kl}^* z_k^* - \gamma' y_{kl}^* s_l^* - \mu_3 y_{kl}^*, \quad \text{for } k = 1, \ldots, m, \text{ and } l = 1, \ldots, n, \tag{A.47c}$$

$$0 = \nu \delta_k \sum_{j=1}^{n} y_{kj}^* - \mu_4 z_k^*, \quad \text{for } k = 1, \ldots, m, \tag{A.47d}$$

$$0 = \nu' \sum_{i=1}^{m} (1 - \delta_i) y_{il}^* - \mu_5 s_l^*, \quad \text{for } l = 1, \ldots, n. \tag{A.47e}$$

Using the notation $\phi_{\text{Tax}} = \nu/\mu_4$ and $\phi_{\text{HBZ}} = \nu'/\mu_5$ to denote the respective growth/death ratios of Tax- and HBZ-specific CTLs as defined in (4.3), we manipulate the equations for z_k^*, s_l^*, and u_{kl}^* and write them in terms of the y_{ij}^*'s. In particular, by Equation (A.47d),

$$z_k^* = \phi_{\text{Tax}} \delta_k \sum_{j=1}^{n} y_{kj}^*, \quad \text{for } k = 1, \ldots, m, \tag{A.48}$$

and by Equation (A.47e),

$$s_l^* = \phi_{\text{HBZ}} \sum_{i=1}^{m} (1 - \delta_i) y_{il}^*, \quad \text{for } l = 1, \ldots, n. \tag{A.49}$$

In addition, Equation (A.47c) yields

$$u_{kl}^* = \frac{1}{\tau_{kl}} \Big[\gamma z_k^* + \gamma' s_l^* + \mu_3 \Big] y_{kl}^*, \quad \text{for } k = 1, \ldots, m, \text{ and } l = 1, \ldots, n. \tag{A.50}$$

Then, substituting Equations (A.48) and (A.49) into Equation (A.50) gives us that for each $k = 1, \ldots, m$ and $l = 1, \ldots, n$,

$$u_{kl}^* = \frac{1}{\tau_{kl}} \left[\gamma \phi_{\text{Tax}} \delta_k \sum_{j=1}^{n} y_{kj}^* + \gamma' \phi_{\text{HBZ}} \sum_{i=1}^{m} (1 - \delta_i) y_{il}^* + \mu_3 \right] y_{kl}^*. \tag{A.51}$$

Next, by Equation (A.47b), we have that

$$u_{kl}^* = \frac{1}{\tau_{kl} + \mu_2} \big(\beta x^* + r \big) y_{kl}^*, \quad \text{for } k = 1, \ldots, m, \text{ and } l = 1, \ldots, n. \tag{A.52}$$

Equating the expressions in (A.51) and (A.52) leads to the following relationship between x^* and the various y_{ij}^*'s. Namely, for $k = 1, \ldots, m$ and $l = 1, \ldots, n$,

$$\mu_3 \big[R_{kl}(x^*) - 1 \big] = \gamma \phi_{\text{Tax}} \delta_k \sum_{j=1}^{n} y_{kj}^* + \gamma' \phi_{\text{HBZ}} \sum_{i=1}^{m} (1 - \delta_i) y_{il}^*, \tag{A.53}$$

where

$$R_{kl}(x) = \frac{\tau_{kl}}{\mu_3(\tau_{kl}+\mu_2)}\big(\beta x + r\big).$$

When all strains have identical fitness, the basic reproduction numbers $R_{0,kl}$ for each strain are equal, which is equivalent to the parameter restriction that $\tau_{kl} = \tau$ for all $k = 1, \ldots, m$ and $l = 1, \ldots, n$. Then,

$$R_{kl}(x) = R(x) = \frac{\tau}{\mu_3(\tau+\mu_2)}\big(\beta x + r\big), \quad \text{for all } k = 1, \ldots, m \text{ and } l = 1, \ldots, n, \tag{A.54}$$

and hence equating the relation (A.53) for any two Tax variants $k_1, k_2 \in \{1, \ldots, m\}$ yields

$$\delta_{k_1} \sum_{j=1}^{n} y^*_{k_1 j} = \delta_{k_2} \sum_{j=1}^{n} y^*_{k_2 j}.$$

In particular, by letting $k_2 = 1$, we find that for each $k_1 \in \{1, \ldots, m\}$,

$$\delta_{k_1} \sum_{j=1}^{n} y^*_{k_1 j} = \delta_1 \sum_{j=1}^{n} y^*_{1j}. \tag{A.55}$$

Next, we compare the frequencies of each pool of epitope-specific immune cells to determine CTL dominance. Summing up the Tax- or HBZ-specific CTL populations as described by Equations (A.48) and (A.49), we compute the total abundance of Tax-specific and HBZ-specific CTLs, respectively, to be

$$z^* = \sum_{i=1}^{m} z^*_i = \phi_{\text{Tax}} \sum_{i=1}^{m} \delta_i \sum_{j=1}^{n} y^*_{ij} \quad \text{and} \quad s^* = \phi_{\text{HBZ}} \sum_{i=1}^{m} \sum_{j=1}^{n} (1-\delta_i) y^*_{ij}. \tag{A.56}$$

According to Equation (A.55), the cumulative abundance of virus strains that all share the same Tax variant is related to the sum of all virus strains that display Tax variant 1. Applying Equation (A.55) to the equations in (A.56) thus yields

$$z^* = m\phi_{\text{Tax}}\delta_1 \sum_{j=1}^{n} y^*_{1j}, \tag{A.57}$$

and

$$s^* = \phi_{\text{HBZ}}\delta_1 \left(\sum_{j=1}^{n} y^*_{1j} \right) \sum_{i=1}^{m} \frac{(1-\delta_i)}{\delta_i}. \tag{A.58}$$

The two CTL populations are equal in frequency when $z^* = s^*$ or, equivalently,

$$\sum_{i=1}^{m} \frac{1}{\delta_i} = m \left[\frac{\phi_{\text{Tax}}}{\phi_{\text{HBZ}}} + 1 \right].$$

Lastly, if we assume that all CTLs are identical with respect to their effective expansion rates, i.e. $\phi_{\text{Tax}} = \phi_{\text{HBZ}}$, then

$$\sum_{i=1}^{m} \frac{1}{\delta_i} = 2m, \tag{A.59}$$

which demonstrates the general result for equal abundances of Tax-specific and HBZ-specific CTLs when there exist m different antigenic variants in the Tax epitope. In the case with two Tax variants (i.e. $m = 2$), the expression, $1/\delta_1 + 1/\delta_2 = 4$, describes the relationship between the immunogenicities δ_1 and δ_2 of the two Tax variants that result in equal numbers of both types of CTLs. As a final remark, observe that Equation (A.59) makes no assumptions on the efficiencies of the two distinct CTL types.

A.5.2 Statement and Proof of Degeneracy in the Multi-Locus Model

Let us assume that an interior equilibrium $E^* = \big(x^*, u_{kl}^*, y_{kl}^*, z_k^*, s_l^*\big)$, where each entry is positive for all $k = 1, \ldots, m$ and $l = 1, \ldots, n$, exists so that it satisfies the steady state equations (A.47). We will prove the result below for the general multi-locus model (6.1).

Theorem A.3 (Degeneracy in the Multi-Locus HTLV-I Model)**.** *For the multi-locus model* (6.1)*, a unique interior equilibrium, corresponding to the co-existence of all possible strains, cannot exist. Instead, we have the following:*

(i) If all strains have identical fitness (i.e. τ_{kl} are equal for all $k = 1, \ldots, m$ and $l = 1, \ldots, n$), then there exists an infinite number of interior equilibria;

(ii) If at least one strain differs in fitness from the others, then no interior equilibrium exists.

Proof. For each $k = 1, \ldots, m$ and $l = 1, \ldots, n$, we have previously derived the following equations (A.53) at equilibrium,

$$\gamma\phi_{\text{Tax}}\delta_k \sum_{j=1}^{n} y_{kj}^* + \gamma'\phi_{\text{HBZ}} \sum_{i=1}^{m} (1 - \delta_i) y_{il}^* = \mu_3\big[R_{kl}(x^*) - 1\big],$$

which represent the rows of a system of linear equations with respect to the variables y_{kl}^*. In particular, let $\big\{(\mathbf{i})\big\}_{i=1,\ldots,mn} = \big\{kl\big\}_{\substack{k=1,\ldots,m \\ l=1,\ldots,n}}$ denote set of lexicographic orderings[3] of the mn distinct virus strains. Then, we can write the above equations to obtain the following system:

$$\mathbf{M}\mathbf{y}^* = \mathbf{c}, \tag{A.60}$$

[3]Lexicographic ordering refers to the same type of ordering as used in the dictionary.

where $\mathbf{y}^* = \left(y^*_{(\mathbf{i})}\right)^T_{\mathbf{i}=1,\ldots,mn} = \left(y^*_{kl}\right)^T_{\substack{k=1,\ldots,m\\ l=1,\ldots,n}}$ and $\mathbf{c} = \left(\mu_3\big[R_{kl}(x^*)-1\big]\right)^T_{\substack{k=1,\ldots,m\\ l=1,\ldots,n}}$ are vectors in $\mathbb{R}^{mn}$ whose entries are indexed according to lexicographic ordering, and the entries of the $mn \times mn$ square (constant) coefficient matrix $\mathbf{M} = \left[m^{(\mathbf{j})}_{(\mathbf{i})}\right]_{\substack{i=1,\ldots,mn\\ j=1,\ldots,mn}}$ are given as follows: For indices $(\mathbf{i}) = k_1 l_1$ and $(\mathbf{j}) = k_2 l_2$, with $k_1, k_2 \in \{1,\ldots,m\}$ and $l_1, l_2 \in \{1,\ldots,n\}$,

$$m^{(\mathbf{j})}_{(\mathbf{i})} = \begin{cases} \gamma\phi_{\text{Tax}}\delta_k + \gamma'\phi_{\text{HBZ}}(1-\delta_k) & \text{when } k_1 = k_2 \text{ and } l_1 = l_2, \\ & \text{(i.e. along the diagonal),} \\ \gamma\phi_{\text{Tax}}\delta_k & \text{when } k_1 = k_2 \text{ and } l_1 \neq l_2, \\ \gamma'\phi_{\text{HBZ}}(1-\delta_k) & \text{when } k_1 \neq k_2 \text{ and } l_1 = l_2, \\ 0 & \text{when both } k_1 \neq k_2 \text{ and } l_1 \neq l_2. \end{cases} \tag{A.61}$$

For $(\mathbf{i}) = kl$, where $k = 1,\ldots,m$ and $l = 1,\ldots,n$, define

$$g_{(\mathbf{i})}(\mathbf{y}^*) = \gamma\phi_{\text{Tax}}\delta_k \sum_{j=1}^{n} y^*_{kj} + \gamma'\phi_{\text{HBZ}} \sum_{i=1}^{m}(1-\delta_i)y^*_{il}. \tag{A.62}$$

Then, Equation (A.53) (which are the rows of the matrix $\mathbf{M}$) can be written succinctly as

$$g_{(\mathbf{i})}(\mathbf{y}^*) = \mathbf{c}_{(\mathbf{i})}, \quad \text{for each } (\mathbf{i}) \in \{(1),\ldots,(mn)\}. \tag{A.63}$$

Now, consider any two indices $(\mathbf{i_0}) = i_1 i_2$ and $(\mathbf{j_0}) = j_1 j_2$, where $i_1, j_1 \in \{1,\ldots,m\}$ and $i_2, j_2 \in \{1,\ldots,n\}$. Let $(\hat{\mathbf{i}}_\mathbf{0}) = i_1 j_2$ and $(\hat{\mathbf{j}}_\mathbf{0}) = j_1 i_2$. We observe that the sum for index $(\mathbf{i_0})$ as in Equation (A.62) can be written in terms of the three other indices (namely, $(\hat{\mathbf{i}}_\mathbf{0})$, $(\hat{\mathbf{j}}_\mathbf{0})$, and $(\mathbf{j_0})$). Specifically,

$$g_{(\mathbf{i_0})}(\mathbf{y}^*) = g_{(\hat{\mathbf{i}}_\mathbf{0})}(\mathbf{y}^*) + g_{(\hat{\mathbf{j}}_\mathbf{0})}(\mathbf{y}^*) - g_{(\mathbf{j_0})}(\mathbf{y}^*). \tag{A.64}$$

Consequently, the rows of $\mathbf{M}$ are linearly dependent, meaning that $\det(\mathbf{M}) = 0$. In other words, the system defined by the equations in (A.63) for $i = 1,\ldots,mn$ is an underdetermined linear system (in terms of the components y^*_{kl}), and so there are two possibilities with respect to the solution $\mathbf{y}^*$ of $\mathbf{My} = \mathbf{c}$: Either there are infinitely many solutions, or there are no solutions.

Considering the same indices above, then for index $(\mathbf{i_0})$, we have

$$g_{(\mathbf{i_0})}(\mathbf{y}^*) = \mathbf{c}_{(\mathbf{i_0})}. \tag{A.65}$$

Next, since the left-hand side of Equation (A.65) can be written as a sum of the other indices according to relation (A.64), we can write the corresponding right-hand side for index $(\mathbf{i_0})$ in an alternative way as follows:

$$g_{(\mathbf{i_0})}(\mathbf{y}^*) = \mathbf{c}_{(\hat{\mathbf{i}}_\mathbf{0})} + \mathbf{c}_{(\hat{\mathbf{j}}_\mathbf{0})} - \mathbf{c}_{(\mathbf{j_0})}. \tag{A.66}$$

In the first scenario (case (i)), when all strains have identical fitness, $\mathbf{c_{(i)}}$ are equal for all indices $i = 1, \ldots, mn$. As a result, the right-hand sides of Equations (A.65) and (A.66) are identical. In this situation, system (A.60) is consistent and there are an infinite number of solutions, corresponding to chronic infection steady states where all strain types co-exist. By contrast, in the complementary scenario (case (ii)), when at least one strain differs in fitness from the others, the right-hand sides of Equations (A.65) and (A.66) are generally not equal. Hence, in this case, no (strictly) interior chronic infection equilibrium point is possible, i.e. the co-existence of all strain types at the same time is not possible. Taken together, the above two possibilities exclude the existence of a unique interior equilibrium, thereby completing the proof.

□

Appendix B

Lemmas

In this section, we state and prove an important result that helps us in the construction of Lyapunov functions that we have used to determine the global dynamics of our baseline mathematical model (3.1), the Tax/HBZ model (4.1), as well as the general multiple epitope model (5.1).

Lemma B.1. *Let $\hat{x}$ be a positive constant. The function $f(x) = x - \hat{x} - \hat{x}\log\frac{x}{\hat{x}}$ is positive definite.*

Proof. The function $f(x)$ satisfies $f(\hat{x}) = 0$. From the first and second derivatives, we observe that

$$f'(x) = 1 - \frac{\hat{x}}{x} \quad \text{and} \quad f'(x) = 0 \iff x = \hat{x},$$

and

$$f''(x) = \frac{\hat{x}}{x^2} > 0, \quad \forall x \neq 0, \quad \text{i.e. the graph of } f(x) \text{ is convex.}$$

Hence, the function $f(x)$ is positive except at the point $x = \hat{x}$, at which it attains its minimum value 0. In other words, $f(x)$ is positive definite. The graph of $f(x)$ when $\hat{x} = 1$ is shown in Figure B.1. □

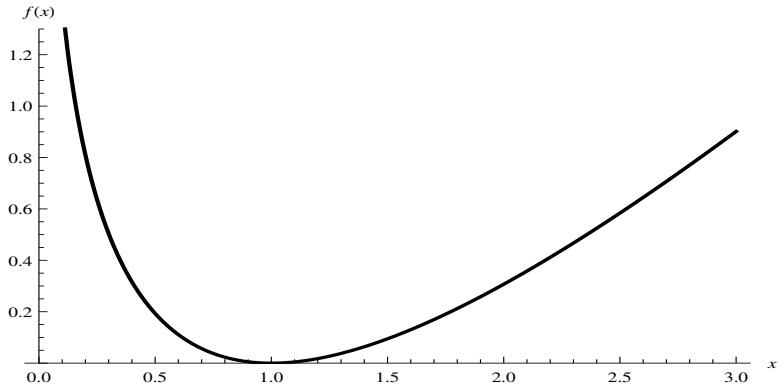


Figure B.1: The graph of the postive definite function $f(x) = x - 1 - \log(x)$.

Appendix C

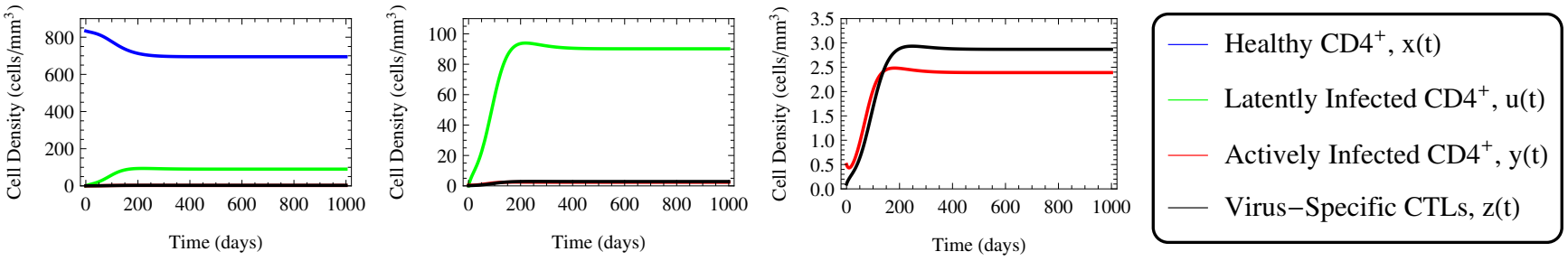
Additional Figures and Tables

C.1 Comparison of Exponential and Logistic Growth Profiles for the Proliferation of Actively Infected Target Cells

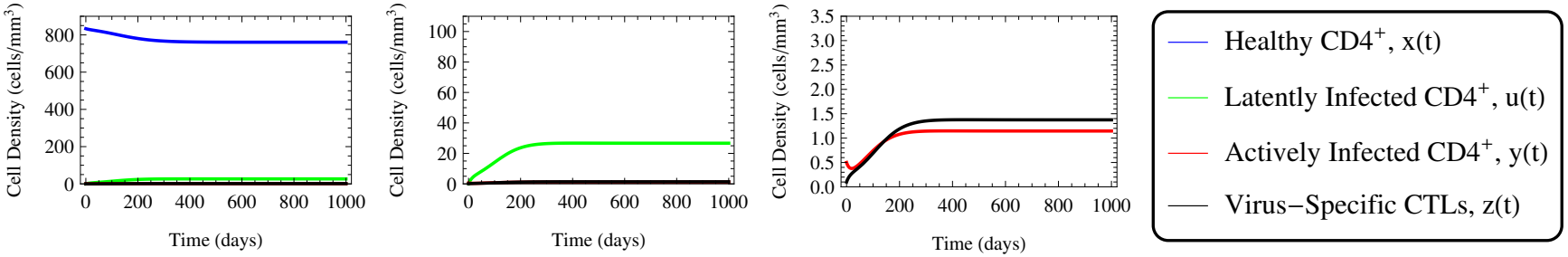
In order to justify the use of an exponential growth profile for the proliferation of provirus-expression infected target cells, we explore the baseline model (3.1) using a full logistic growth term, $ry(t)\big(1-\big(x(t)+u(t)+y(t)\big)/k\big)$, with the biologically realistic parameter values in Table 3.1. In the numerical simulations that follow for the model with logistic growth, we have selected a carrying capacity of $k = 800$ cells/mm^3, which is on the same order as the value for k listed in Table 3.1. We note, however, that our conclusions here are not dependent on any particular choice of k. Next, because the rate r of infected target cell proliferation differentially influences each of the two models, we explore its effect on the various cell populations. We find that when the value of r was chosen in the physiologically relevant range as in Table 3.1, the two models converge to equilibrium points that differ little quantitatively (result not shown). This observation is not unexpected, as we have demonstrated in Section 3.5 using non-dimensionalisation and asymptotic analysis that the impact of the mitotic growth term is small, namely of $O(\epsilon)$, where $\epsilon = 10^{-2}$, at least in the range for r in Table 3.1. We therefore select a value of r that is two orders of magnitude higher than the original value. With respect to the asymptotic analysis in Section 3.5, this amounts to re-scaling the infected cell proliferation term, in both the model with exponential growth and the one with logistic growth, to leading order ($O(1)$).

The time series plots in this case (i.e. when the value for r is large) for each of the T-cell compartments in the two models are shown in Figure C.1 and those of the proviral load and active proportion are shown in Figure C.2 (see Section 3.6 for a definition of proviral load and active proportion with respect to model (3.1)). Although there are quantitative differences between the model with exponential growth versus the model with logistic growth, we observe no qualitative difference in the behaviour of solution trajectories. The primary

effect of a logistic growth term serves to control the magnitude of the infected target cell populations and hence would result in a lower proviral load.



(a) Model with a linear growth term (large r).



(b) Model with a logistic growth term (large r).

Figure C.1: Time series plots for each of the T-cell populations under investigation comparing the dynamical behaviour of the baseline model (3.1) when infected target cell proliferation is described by (a) a linear growth term, and (b) a logistic growth term. Because we noticed that the mitotic growth term, whether exponential or logistic, had a negligible effect on the levels of the various T-cell populations when the rate r of infected target cell proliferation was chosen in the physiologically realistic range as in Table 3.1, to emphasise the role of mitotic transmission, here we have selected a value of r that is two orders of magnitude higher (which we refer to as 'large r'). Morever, for the latter model, the carrying capacity k has been selected so that $x+u+y$ is of the order k. Specifically, $k = 800$ cells/mm^3. All other parameters have been chosen as in Table 3.1 and are the same between the two models. We observe that although the effect of a logistic growth term serves to control the magnitude of each of the infected cell populations, there is no qualitative difference in dynamical behaviour between the two models given by the different growth terms.

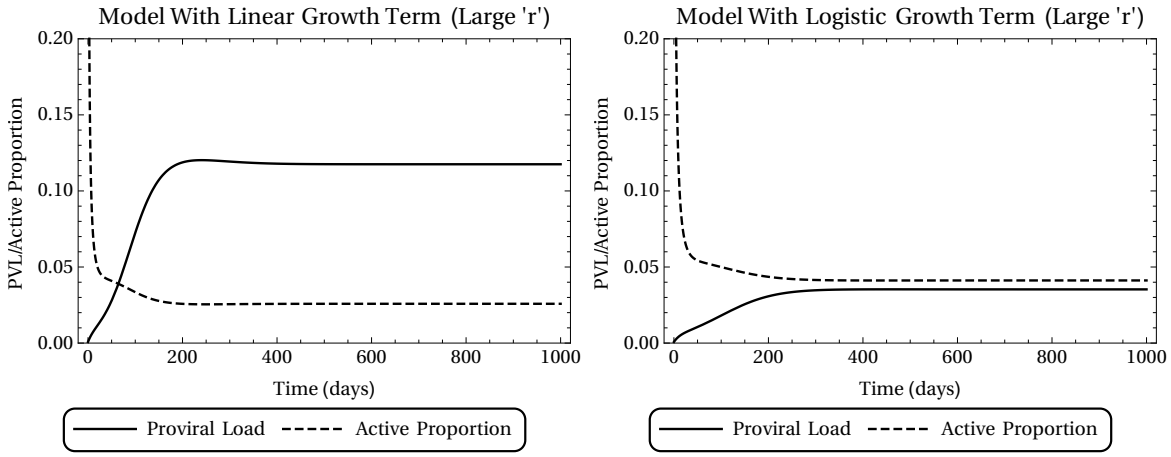


Figure C.2: Time series graphs for the proviral load and active proportion of infected cells comparing linear and logistic growth terms for infected target cell proliferation. In the above figures, as with Figure C.1, the value of r is selected to be two orders of magnitude higher than in Table 3.1 (which we refer to as 'large r') in order to emphasise the role of mitotic proviral transmission for the two models. We observe, as would be expected, that the model with a logistic growth term (right plot) achieves better control of the proviral load than the model with an exponential growth term (left plot). All other parameters have been chosen as in Table 3.1 and are equal for both models. Solid line: Proviral load. Dashed line: Active proportion of infected target cells.

C.2 Original Tables of Experimental Data Used in the Multiple Epitope Model

Table 1. Immunodominance hierarchy of the human T cell lymphotropic virus type (HTLV)–1–specific CD8⁺ T cell response.

Group, subject	Age, years	Provirus load, %	Peptide									
			Tax13	Tax20	Env	Rex	Tof	Rof	Gag	Pol	Pro	Total
HAM/SP												
TAY	78	14.00	8480	ND	5590	0	1540	0	3000	7880	0	26,500
TBA	63	10.90	631	496	2000	0	996	673	0	902	0	5220
TAQ	75	1.50	584	ND	0	0	0	0	0	0	0	584
TAU	73	2.70	534	ND	300	152	243	265	117	2350	0	3961
TBI	68	32.00	3390	ND	0	0	0	0	0	1230	0	4620
TW	38	23.00	6620	3450	980	0	0	0	392	0	0	7990
TBK	78	35.20	4900	2390	2060	0	0	0	416	537	416	8330
TAN	47	6.20	29,000	34,000	3300	0	0	0	0	0	0	37200
TAE	80	3.10	7850	8220	337	0	214	241	228	296	0	9170
TAL	66	26.20	4460	3160	617	0	0	0	225	0	0	5300
AC												
HBD	49	0.01	8130	6430	0	0	0	0	394	947	0	9470
HBE	70	11.50	8960	6160	2170	0	0	563	2290	883	374	15000
HBF	37	6.50	857	690	0	0	0	290	273	623	0	2040
HT	75	1.20	15,900	5300	1680	0	0	0	0	4940	0	22520
HAY	61	10.60	4110	0	0	0	0	0	0	0	1570	5680
HX	60	0.01	4670	2960	772	0	2320	822	1810	4530	0	14924
HAE	57	0.01	0	0	0	0	0	0	0	5690	0	5690
Un												
Un1	40	NA	0	0	0	0	0	0	0	0	0	0
Un2	48	NA	0	0	0	0	0	0	0	0	0	0
Un3	65	NA	0	0	0	0	0	0	0	0	0	0

NOTE. Patients with HTLV-1–associated myelopathy/tropical spastic paraparesis (HAM/TSP) have anonymized codes beginning with T, asymptomatic HTLV-1 carriers (ACs) have codes beginning with H, and uninfected control subjects have codes beginning with Un. All results shown are the mean of replicate well counts minus the mean negative (background) and are expressed in no. of spot-forming cells/10^6 CD8⁺ T cells. Counts were considered to be positive if values were ⩾(2 SD + mean negative [mean of a minimum of 6 negative wells]). All non-Tax peptides are 20mers. 0, below the limit of detection; NA, not applicable; ND, not done; Tax13, Tax 13mer peptides (overlapping by 9); Tax20, Tax 20mer peptides (overlapping by 14).

Supplementary Table C.1: Original table of experimental data from Goon et al. (2004a) (Table 1). The red rectangle indicates the general portion of the data set which we have used in our model (refer to Table 5.6). (Figure reproduced with permission from the publisher.)

Table 1. Subject data and $CD8^+$ cell antiviral efficacy estimates

Disease duration is measured in years since first diagnosis. $CD8^+$ cell antiviral efficacy is defined as the proportion of Tax^+CD4^+ cells cleared per $CD8^+$ cell per day. Negative estimates of antiviral efficacy were likely to be due to noise contributing in such a way that near zero clearance rates appeared negative. A-C, Afro-Caribbean; Cauc, caucasian; ND, not done; Dis. Dur., disease duration.

Subject	Clinical status	Sex	Ethnicity	Dis. dur. (years)	Age (years)	Proviral load (% PBMC)	Antiviral efficacy (per $CD8^+$ cell per day)		
							Repeat 1	Repeat 2	Mean
TAC	HAM/TSP	F	A-C	16	56	12·1	0·088	0·093	0·091
TAY	HAM/TSP	F	A-C	8	60	2·2	0·317	0·279	0·298
TBA	HAM/TSP	F	A-C	18	61	3·5	0·060	0·040	0·050
TAT	HAM/TSP	F	A-C	8	66	4·2	0·067	0·050	0·058
TBG	HAM/TSP	F	A-C	7	66	16·4	0·010	0·005	0·007
TW	HAM/TSP	F	A-C	10	36	10·3	0·039	0·009	0·024
TAQ	HAM/TSP	M	A-C	18	73	1·0	0·081	0·084	0·083
TAU	HAM/TSP	M	A-C	11	72	5·4	0·051	0·047	0·049
TBI	HAM/TSP	M	A-C	11	66	12·3	−0·002	0·007	0·003
TAL	HAM/TSP	F	A-C	26	65	ND	0·079	0·026	0·053
TAA	HAM/TSP	F	A-C	17	64	ND	0·051	0·045	0·048
TBJ	HAM/TSP	F	A-C	13	42	ND	0·045	0·034	0·039
TBO	HAM/TSP	F	A-C	9	56	ND	0·198	0·145	0·172
TAZ	HAM/TSP	M	Asian	22	64	ND	0·015	0·024	0·020
HS	AC	F	A-C	–	47	1·6	−0·029	0·002	−0·013
HS	AC	F	A-C	–	47	5·8	0·003	−0·001	0·001
HT	AC	F	A-C	–	74	1·0	0·073	0·050	0·062
HY	AC	F	A-C	–	58	0·4	0·094	0·036	0·065
HAY	AC	F	Persian	–	60	10·6	−0·001	−0·012	−0·007
HBD	AC	F	Cauc.	–	47	0·001	0·182	0·258	0·220
HBF	AC	F	A-C	–	56	2·8	0·033	0·024	0·029
HBH	AC	F	A-C	–	35	3·98	0·018	0·021	0·020
HCB	AC	F	A-C/Cauc.	–	39	ND	0·038	0·064	0·051
HBE	AC	F	A-C	–	69	ND	0·362	0·492	0·427

Supplementary Table C.2: Experimental data from Asquith et al. (2005a) (Table 1). The red rectangle indicates the general portion of the data set which we take for parameter estimation using our multiple epitope model framework in Section 5.4.3. (Figure reproduced with permission from the publisher.)

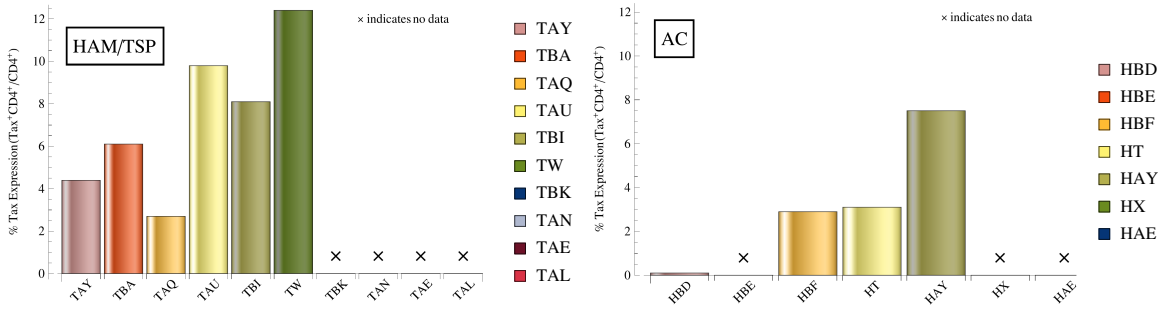
Table 1: Tax expression in CD4⁺ lymphocytes is 2.5–3 fold higher in HAM/TSP patients than ACs of comparable proviral load

		Patient	Proviral load (% PBMC)	Rate of CTL lysis (per CD8+ cell per day)	%Tax expression (Tax+CD4+/CD4+)	Mean Tax expression	Fold Increase in Tax Expression (HAM÷AC)
Group 1	AC	HBD	0.0	0.220	0.1	1.3	2.8
		HT	1.0	0.062	3.1		
		HY	0.4	0.065	0.6		
	HAM	TAQ	1.0	0.083	2.7	3.5	
		TAY	2.2	0.298	4.4		
Group 2	AC	HBH	4.0	0.020	2.7	2.8	3.0
		HBF	2.8	0.029	2.9		
	HAM	TAT	4.2	0.058	9.6	8.5	
		TAU	5.4	0.049	9.8		
		TBA	3.5	0.050	6.1		
Group 3	AC	HS	5.8	0.001	3.0	5.2	2.5
		HAY	10.6	-0.007	7.5		
	HAM	TW	10.3	0.024	12.4	13.3	
		TAC	12.1	0.091	11.1		
		TBG	16.4	0.007	21.5		
		TBI	12.3	0.003	8.1		

The choice of groups of "comparable" proviral load is, to some extent, subjective but a range of alternative groupings gave similar results. This included a grouping in which the mean proviral load of ACs was higher than the mean proviral load of HAM/TSP patients in each group (it was necessary to omit some high proviral load HAM/TSP patients in order to obtain this alternative grouping). We have illustrated our results using this particular grouping because it is a representative grouping and because it yields two or more subjects in each group thus minimising the effect of outliers.

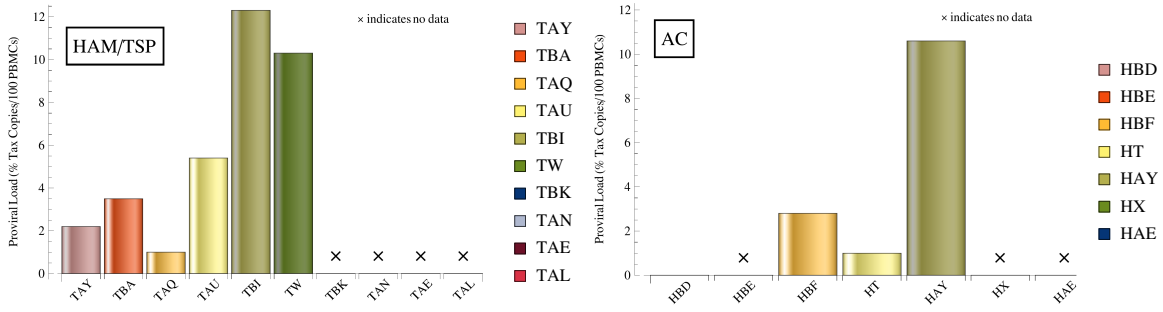
Supplementary Table C.3: Experimental data from Asquith et al. (2005b) (Table 1). The red rectangle indicates the general portion of the data set which we take for parameter estimation using our multiple epitope model framework in Section 5.4.3. (Figure reproduced with permission from the publisher.)

C.3 Additional Bar Charts and Combined Plots for Parameter Estimation Using the Multiple Epitope Model



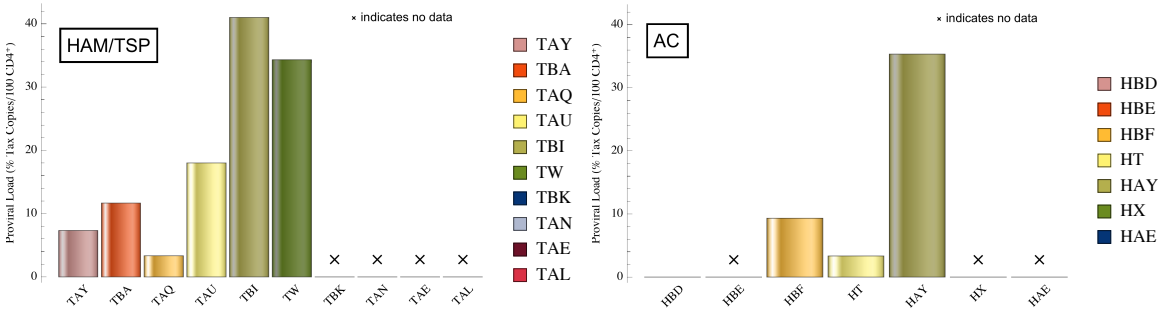
(a) Measured Tax expression in HAM/TSP patients.

(b) Measured Tax expression in ACs.



(c) Measured PVL in HAM/TSP patients.

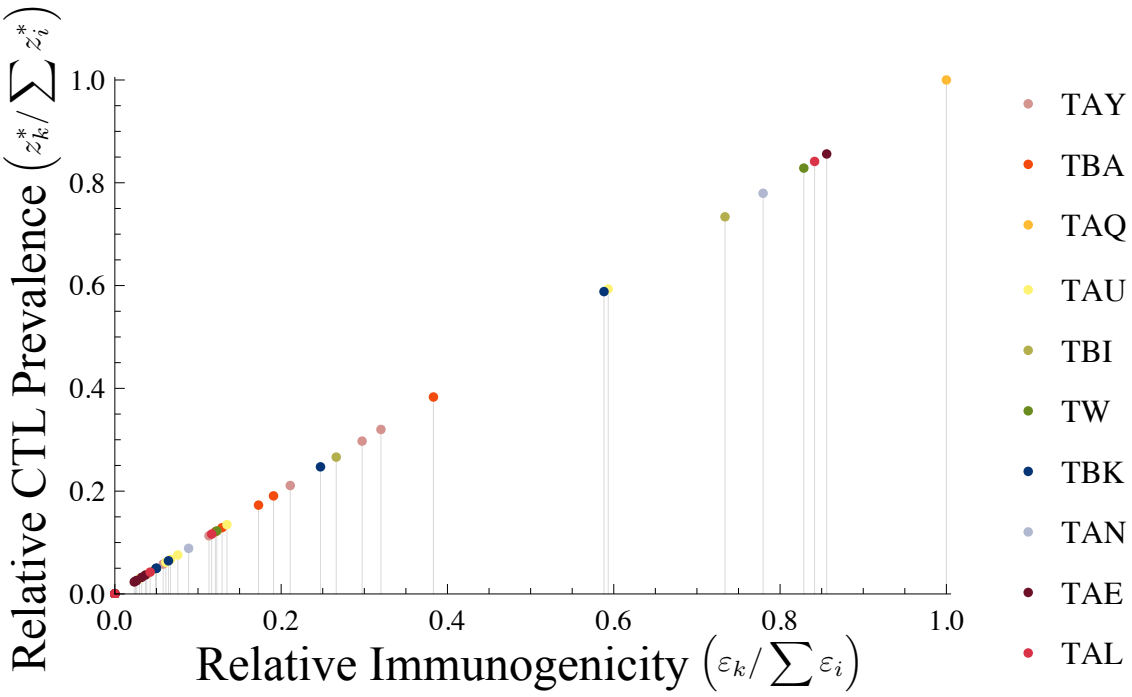
(d) Measured PVL in ACs.



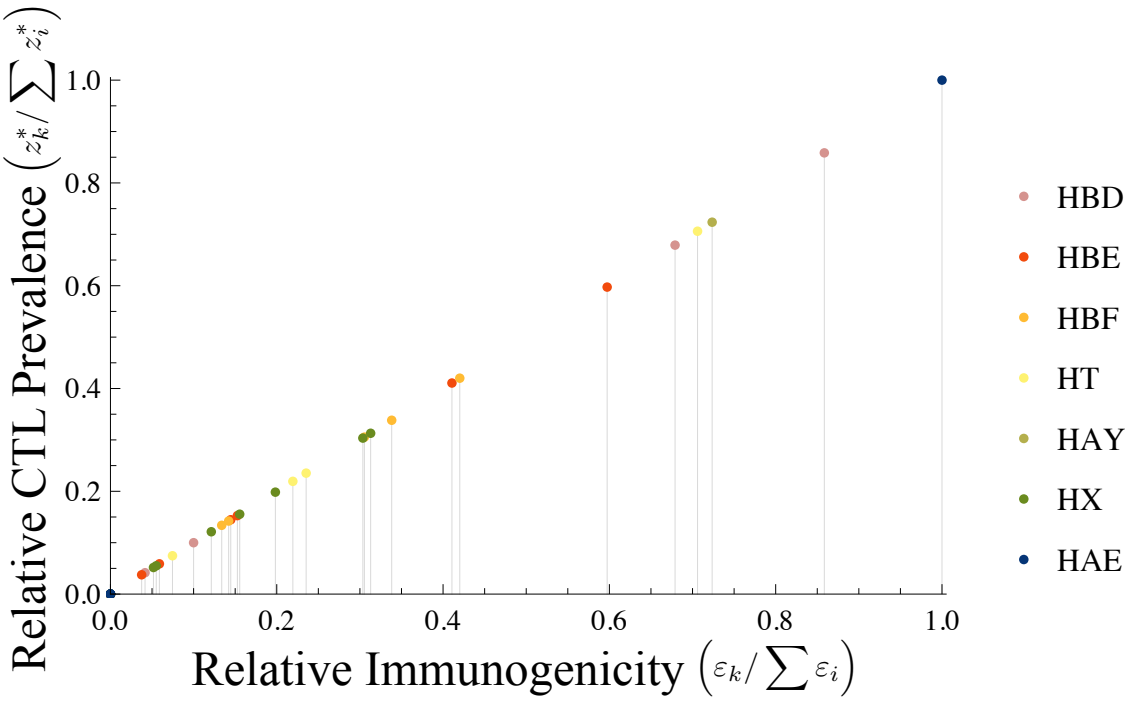
(e) Adjusted PVL in HAM/TSP patients.

(f) Adjusted PVL in ACs.

Figure C.4: Bar chart representation describing the experimental data from Asquith et al. (2005a,b). (a)–(b) Measurements of the Tax$^+$ pool of CD4$^+$ T-cells. Refer to Figure 5.9(a) for a bar chart combining the data from both HAM/TSP patients and ACs. (c)–(d) Measurements of the proviral load in terms of the percentage of Tax copies per 100 PBMCs. (e)–(f) Adjusted measurements for the proviral load in terms of the percentage of Tax copies per 100 CD4$^+$ helper T-cells. These adjusted values were obtained by dividing the measured PVL with respect to the pool of PBMCs by the scaling factor of 0.3 CD4$^+$/PBMC as described in Subsection 5.4.2. Refer to Figure 5.9(b) for a bar chart combining the data from both HAM/TSP patients and ACs. Refer to Table 5.7, shaded rows, for the values. ×: indicates no data available for the particular individual.

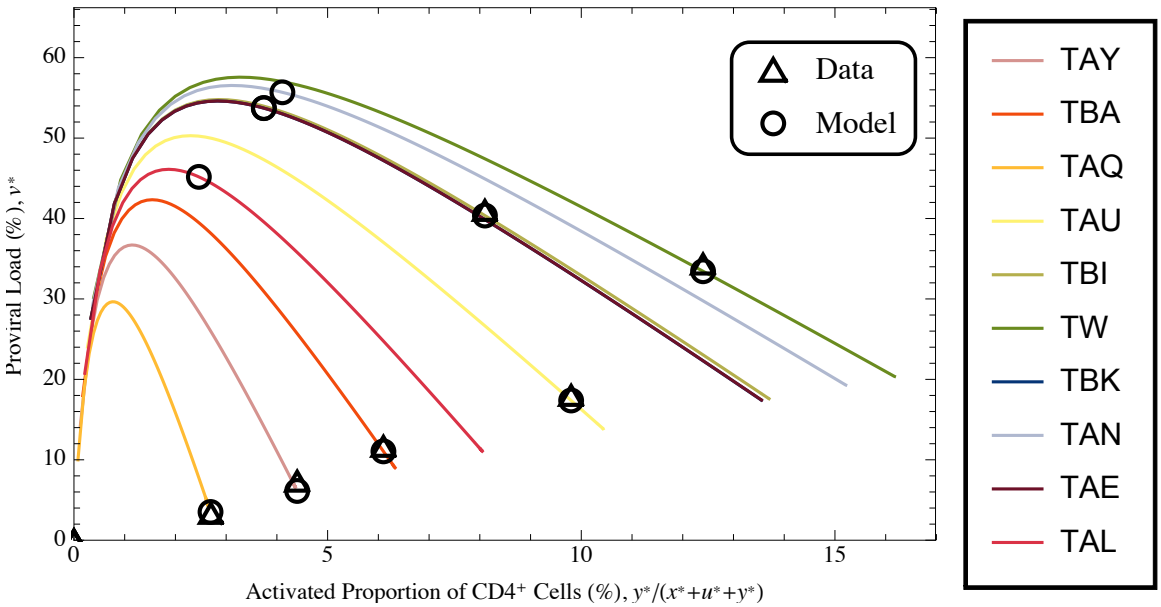


(a) Combined immunogenicity profiles for HAM/TSP patients in the data.

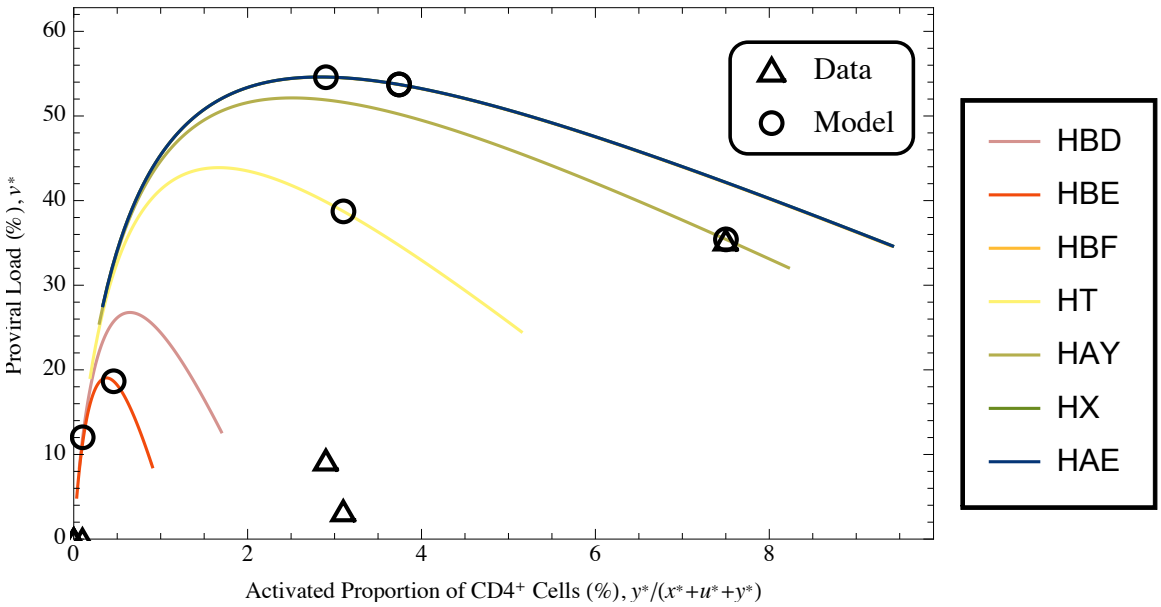


(b) Combined immunogenicity profiles for ACs in the data.

Figure C.5: Combined list plot representation of the immunogenicity profiles of HTLV-I-infected individuals as described by experimentally measured data from Goon et al. (2004a) for (a) 10 HAM/TSP patients, and (b) 7 ACs. The figure shows the distribution of relative immunogenicities towards the eight epitopes (horizontal axis) and the relative abundances of their cognate CTL types (vertical axis) for the 17 HTLV-I seropositive subjects in each of the two infected cohorts. The individual immunogenicity profiles for each subject is shown in Figure 5.10.



(a) Combined plots displaying viral burden and parameter estimation in the subset of HAM/TSP patients.



(b) Combined plots displaying viral burden and parameter estimation in the subset of ACs.

Figure C.6: Combined plots showing the matching of the multiple epitope model to experimental data, which allows us to derive estimates for two key parameters in each of the two HTLV-I seropositive cohorts, namely, for (a) 10 HAM/TSP patients, and (b) 7 ACs. The measured and estimated values for the various parameters and quantities related to viral detriment is shown in Table 5.9. △ Experimental data points. ◯ Estimated point using model (5.1). Solid line: Curve describing viral detriment as a parametric function of τ, the rate of spontaneous proviral expression, as predicted by the multiple epitope model (5.1). The individual plots for each infected subject are shown in Figure 5.11.

C.4 Strain Structure in an Asymmetric Multi-Locus System: The 3×2 Model

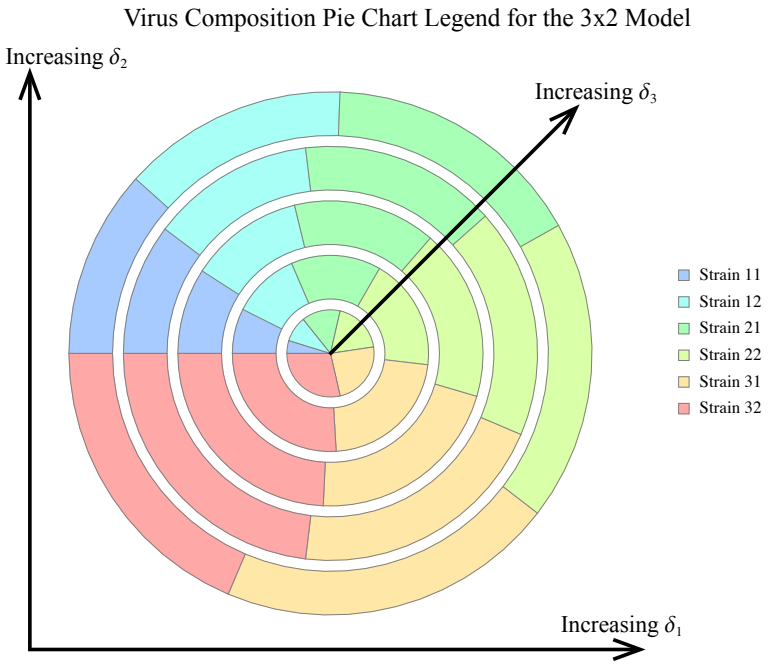


Figure C.7: Pie chart legend indicating virus composition in terms of relative abundances of each individual viral strain for asymmetric multi-locus model with three variants in Tax and two variants in HBZ. This schematic legend illustrates a 'concentric pie chart representation' to visualise the structure of the virus population at equilibrium according to model (6.1) in the case $m = 3$ and $n = 2$, i.e. the 3×2 model (c.f. Figure 6.13, which shows the same type of representation for the 3×3 model). We therefore consider a total of nine possible virus strains which, listed in lexicographic order, are: 11, 12, 21, 22, 31 and 32. In this representation, we consider a series of circular plots (like the one shown above) consisting of a set of concentric pie charts. Each individual pie chart corresponds to a particular combination of selected values for the immunogenicities of each of the three distinct Tax variants; that is, δ_1, δ_2 and δ_3. In Figures C.8–C.10, we consider five different sample choices, representing a set of evenly spaced values between 0 and 1, for each of the δ_i. Separate circular plots along the horizontal and vertical axes correspond to the various choices of δ_1 and δ_2, respectively (note that the schematic legend above only shows one such circular plot). Meanwhile, within each circular plot, the pie charts spreading radially outward correspond to increasing values of δ_3. For each pie chart, the colours represent individual viral strains as indicated.

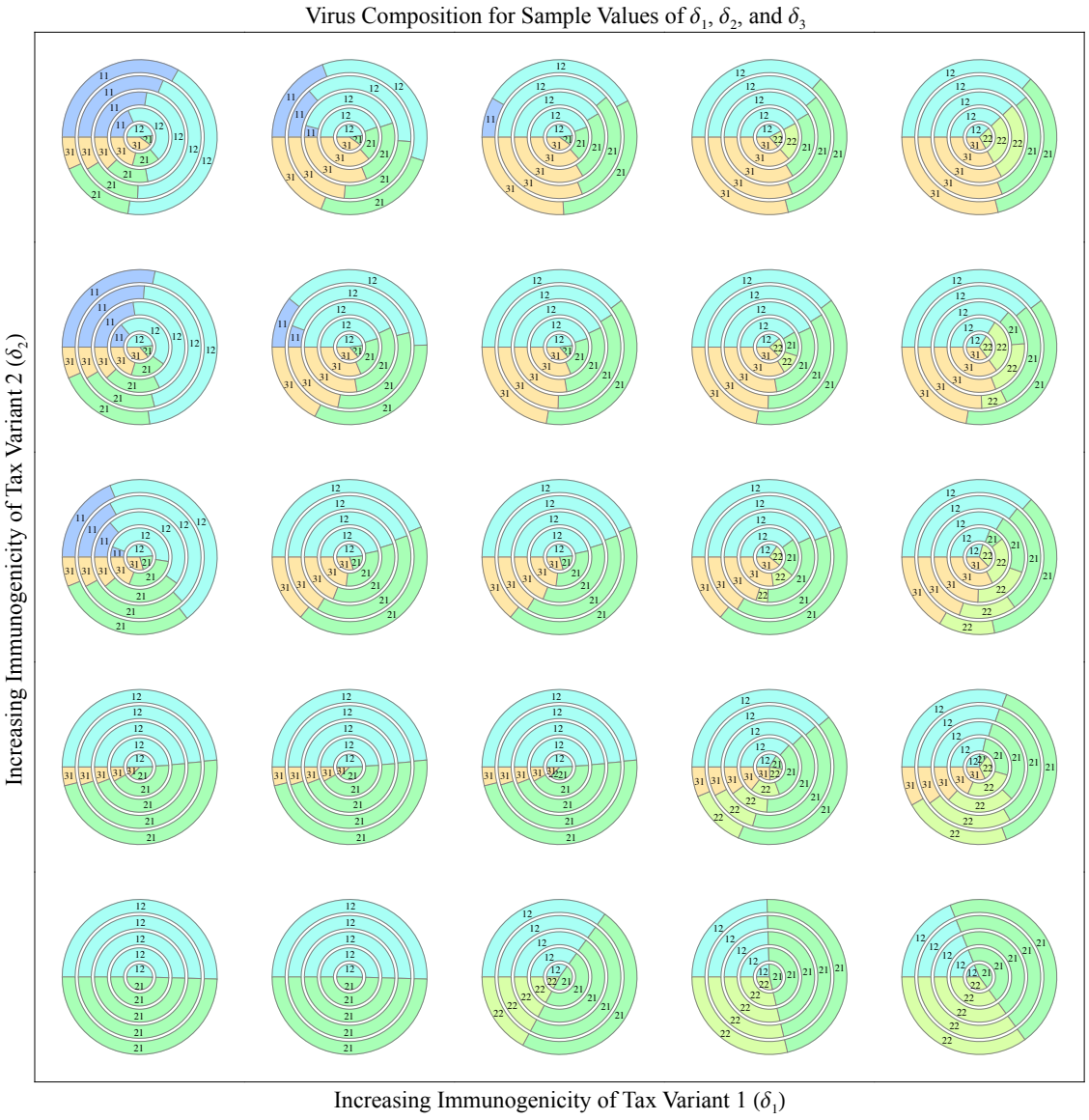


Figure C.8: Visualisation of virus composition at equilibrium in the asymmetric 3×2 multi-locus model. The above figure shows one possible outcome of solving model (6.1) with $m = 3$ and $n = 2$ numerically, and plotting the relative abundances of all co-existing virus strains at equilibrium for a particular combination of Tax variant immunogenicities (i.e. δ_1, δ_2, and δ_3). Strains ij are labelled as indicated. We observe that a wide variety of overlapping and non-overlapping strain structures are possible. In this figure, the basic reproduction numbers for each strain are $R_{0,11} = 7.70$, $R_{0,12} = 9.95$, $R_{0,21} = 9.71$, $R_{0,22} = 7.55$, $R_{0,31} = 8.09$, $R_{0,32} = 6.33$. Meanwhile, each of the Tax variants can take on one of the following five (uniformly-spaced) immunogenicity values between 0 and 1: 0.05, 0.275, 0.5, 0.725, 0.95. Thus, a total of $5 \times 5 \times 5 = 125$ simulations are represented. All other parameters have been selected from Table 3.1.

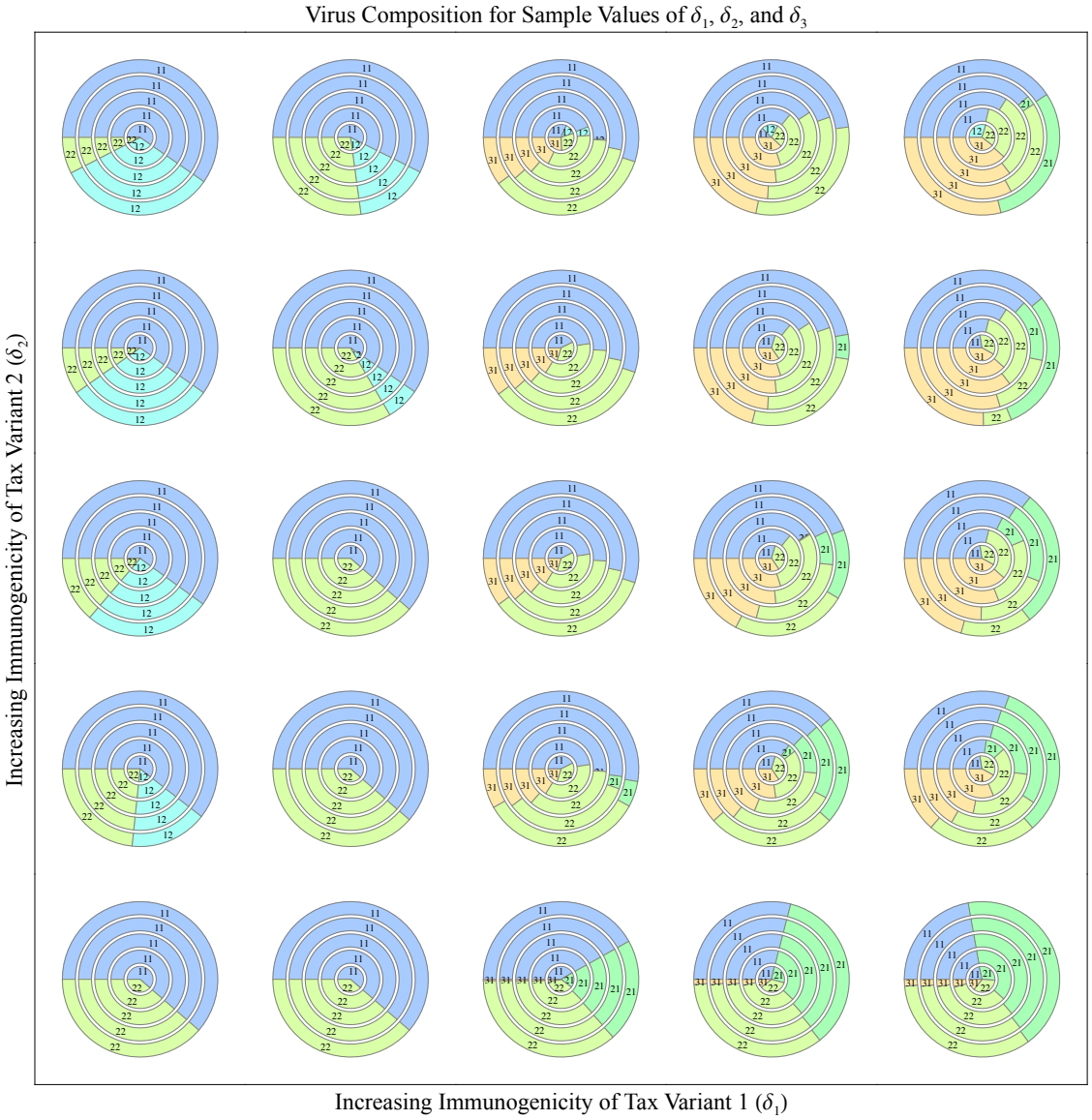


Figure C.9: DSS and strain dominance in the asymmetric 3×2 system. The above figure shows the result of numerical simulation of the multi-locus model (6.1) with $m = 3$ and $n = 2$. When there is a set of discordant strains, the strain with the highest fitness often dominates the viral population, i.e. it exists at the highest prevalence. For example, for the parameters considered here, Strain 11 is the fittest strain and co-exists with a less fit discordant strain 22 sharing no common variants at either of the two epitopes. This is shown notably in the concentric pie charts occupying roughly the bottom left portion of the figure, where we observe that when the two strains maintain DSS, Strain 11 is dominant. The fitnesses for the various strains are $R_{0,11} = 11.09$, $R_{0,12} = 6.16$, $R_{0,21} = 8.14$, $R_{0,22} = 6.38$, $R_{0,31} = 7.80$, $R_{0,32} = 5.98$. Each of δ_1, δ_2, and δ_3 can take one of the following five (uniformly-spaced) immunogenicity values between 0 and 1: 0.05, 0.275, 0.5, 0.725, 0.95. Thus, the figure represents a total of $5 \times 5 \times 5 = 125$ model simulations. All parameter values have been selected from physiologically realistic ranges as discussed in Table 3.1.

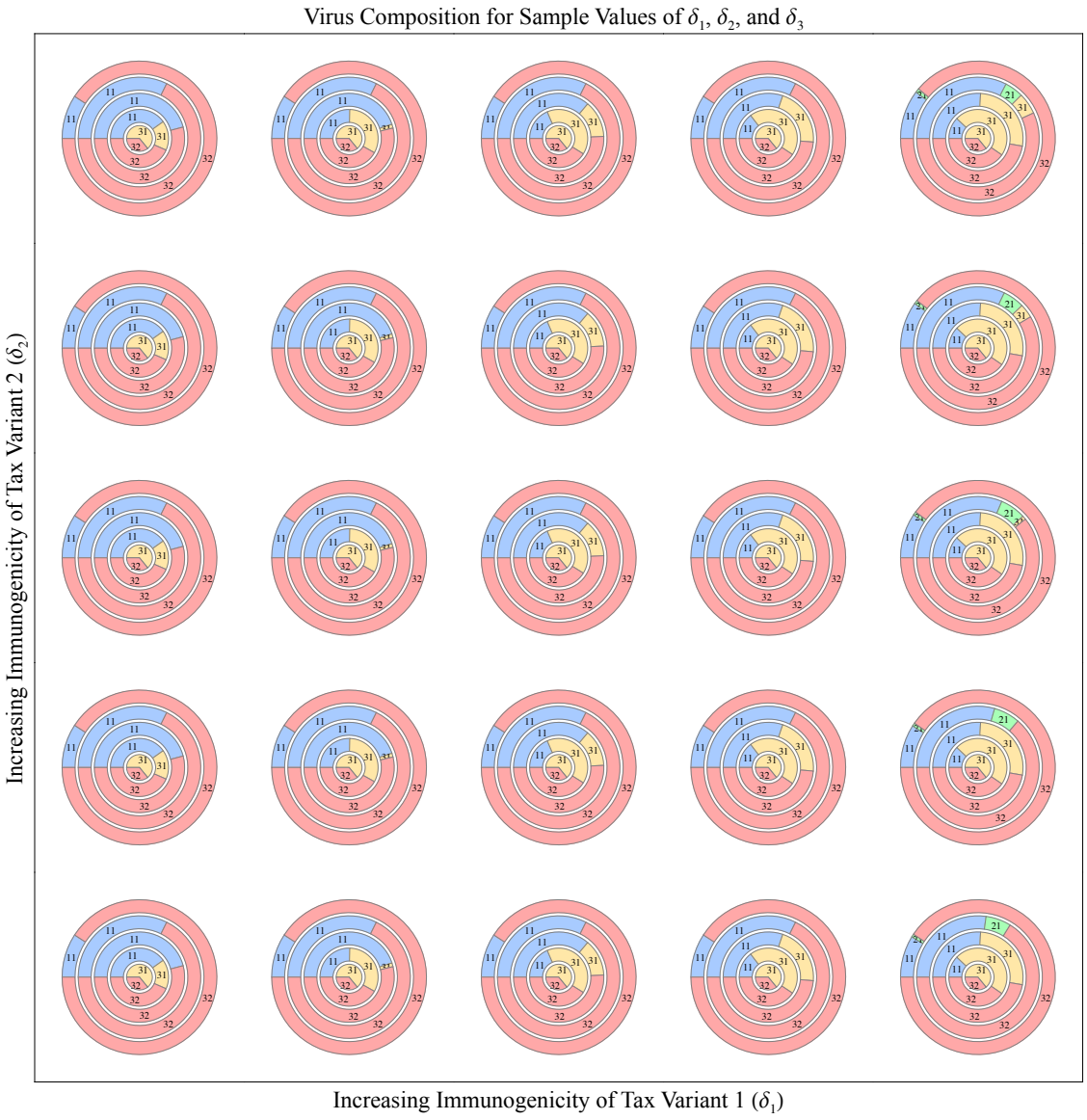


Figure C.10: As with the symmetric multi-locus system, DSS does not imply that the fittest strain is dominant in the asymmetric system. The above figure illustrates yet another possible outcome from numerical exploration of the multi-locus model (6.1) with $m = 3$ and $n = 2$. In this scenario, Strain 31 is the fittest strain, yet for many immunogenicity values for the three Tax variants δ_1, δ_2 and δ_3, we observe instead the co-existence of Strain 11 and Strain 32 forming a discordant, non-overlapping set of strains. Moreover, we notice that the much less fit of the two strains, Strain 32, strongly dominates the virus population, in several instances making up over 90% of provirus-positive cells. Here, $R_{0,11} = 8.87$, $R_{0,12} = 4.29$, $R_{0,21} = 3.24$, $R_{0,22} = 1.17$, $R_{0,31} = 10.66$, $R_{0,32} = 5.27$. Each of the Tax variant immunogenicities can take one of the following five (uniformly-spaced) immunogenicity values: 0.05, 0.275, 0.5, 0.725, 0.95. These are meant to represent the range of values between 0 and 1, and leads to a total of $5 \times 5 \times 5 = 125$ combinations of δ_1, δ_2, and δ_3 (corresponding to each of the 125 individual pie charts illustrated above). All parameter values have been selected from Table 3.1.

Bibliography

A. Akram and R. D. Inman. Immunodominance: a pivotal principle in host response to viral infections. *Clinical Immunology*, 143(2):99–115, 2012.

C. L. Althaus and R. J. D. Boer. Dynamics of immune escape during HIV/SIV infection. *PLoS Computational Biology*, 4(7):e1000103, 2008.

B. Asquith. The evolutionary selective advantage of HIV-1 escape variants and the contribution of escape to the HLA-associated risk of AIDS progression. *PLoS One*, 3(10): e3486, 2008.

B. Asquith and C. R. M. Bangham. The role of cytotoxic T lymphocytes in human T-cell lymphotropic virus type 1 infection. *Journal of Theoretical Biology*, 207(1):65–79, 2000.

B. Asquith and C. R. M. Bangham. The dynamics of T-cell fratricide: application of a robust approach to mathematical modelling in immunology. *Journal of Theoretical Biology*, 222(1):53–69, 2003a.

B. Asquith and C. R. M. Bangham. An introduction to lymphocyte and viral dynamics: the power and limitations of mathematical analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1525):1651–1657, 2003b.

B. Asquith and C. R. M. Bangham. Quantifying HTLV-I dynamics. *Immunology and Cell Biology*, 85:280–286, 2007.

B. Asquith and C. R. M. Bangham. How does HTLV-I persist despite a strong cell-mediated immune response? *Trends in Immunology*, 29(1):4–11, 2008.

B. Asquith, E. Hanon, G. P. Taylor, and C. R. M. Bangham. Is human T-cell lymphotropic virus type I really silent? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1400):1013–1019, 2000.

B. Asquith, A. J. Mosley, A. Barfield, S. E. F. Marshall, A. Heaps, P. Goon, E. Hanon, Y. Tanaka, G. Taylor, and C. R. M. Bangham. A functional $CD8^+$ cell assay reveals individual variation in $CD8^+$ cell antiviral efficacy and explains differences in human T-lymphotropic virus type 1 proviral load. *Journal of General Virology*, 86:1515–1523, 2005a.

B. Asquith, A. J. Mosley, A. Heaps, Y. Tanaka, G. Taylor, A. R. McLean, and C. R. M. Bangham. Quantification of the virus-host interaction in human T lymphotropic virus type I infection. *Retrovirology*, 2(1):75–83, 2005b.

B. Asquith, C. T. T. Edwards, M. Lipsitch, and A. R. McLean. Inefficient cytotoxic T lymphocyte–mediated killing of HIV-1–infected cells in vivo. *PLoS Biology*, 4(4):e90, 2006.

B. Asquith, Y. Zhang, A. J. Mosley, C. M. de Lara, D. L. Wallace, A. Worth, L. Kaftantzi, K. Meekings, G. E. Griffin, Y. Tanaka, D. F. Tough, P. C. Beverly, G. P. Taylor, D. Macallan, and C. R. M. Bangham. *In vivo* T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection. *Proceedings of the National Academy of Sciences*, 104(19):8035–8040, 2007.

I. Bains, R. Thiébaut, A. J. Yates, and R. Callard. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *The Journal of Immunology*, 183(7):4329–4336, 2009.

C. R. M. Bangham. HTLV-1 infections. *Journal of Clinical Pathology*, 53(8):581–586, 2000a.

C. R. M. Bangham. The immune response to HTLV-I. *Current Opinion in Immunology*, 12:397–402, 2000b.

C. R. M. Bangham. The immune control and cell-to-cell spread of human T-lymphotropic virus type 1. *Journal of General Virology*, 84:3177–3189, 2003.

C. R. M. Bangham. CTL quality and the control of human retroviral infections. *European Journal of Immunology*, 39(7):1700–1712, 2009.

C. R. M. Bangham and M. Osame. Cellular immune response to HTLV-1. *Oncogene*, 24: 6035–6046, 2005.

C. R. M. Bangham, A. G. Kermode, S. E. Hall, and S. Daenke. The cytotoxic T-lymphocyte response to HTLV-I: the main determinant of disease? *Seminars in Virology*, 7(1):41–48, 1996.

C. R. M. Bangham, S. E. Hall, K. J. M. Jeffery, A. M. Vine, A. Witkover, M. A. Nowak, D. Wodarz, K. Usuku, and M. Osame. Genetic control and dynamics of the cellular immune response to the human T-cell leukaemia virus, HTLV-I. *Philosophical Transactions: Biological Sciences*, 354(1384):691–700, 1999.

C. R. M. Bangham, K. Meekings, F. Toulza, M. Nejmeddine, E. Majorovits, B. Asquith, and G. Taylor. The immune control of HTLV-I infection: selection forces and dynamics. *Frontiers in Bioscience*, 14:2889–2903, 2009.

W. E. Biddison, R. Kubota, T. Kawanishi, D. D. Taub, W. W. Cruikshank, D. M. Center, E. W. Connor, U. Utz, and S. Jacobson. Human T cell leukemia virus type I (HTLV-I)-specific CD8+ CTL clones from patients with HTLV-I-associated neurologic disease secrete proinflammatory cytokines, chemokines, and matrix metalloproteinase. *The Journal of Immunology*, 159(4):2018–2025, 1997.

K. B. Blyuss and S. Gupta. Stability and bifurcations in a model of antigenic variation in malaria. *Journal of Mathematical Biology*, 58(6):923–937, 2009.

M. Bofill, G. Janossy, C. A. Lee, D. MacDonald-Burns, A. N. Phillips, C. Sabin, A. Timms, M. A. Johnson, and P. B. A. Kernoff. Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. *Clinical and Experimental Immunology*, 88(2):243–252, 1992.

M. Boxus and L. Willems. Mechanisms of HTLV-I persistence and transformation. *British Journal of Cancer*, 101(9):1497–1501, 2009.

C. O. Buckee, K. A. Jolley, M. Recker, B. Penman, P. Kriz, S. Gupta, and M. C. J. Maiden. Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis. *Proceedings of the National Academy of Sciences*, 105(39):15082–15087, 2008.

M. J. Callaghan, C. O. Buckee, K. A. Jolley, P. Kriz, M. C. J. Maiden, and S. Gupta. The effect of immune selection on the structure of the meningococcal opa protein repertoire. *PLoS pathogens*, 4(3):e1000020, 2008.

V. Calvez, A. Korobeinikov, and P. K. Maini. Cluster formation for multi-strain infections with cross-immunity. *Journal of Theoretical Biology*, 233(1):75–83, 2005.

M.-H. Cavanagh, S. Landry, B. Audet, C. Arpin-André, P. Hivin, M.-È. Paré, J. Thête, É. Wattel, S. J. Marriott, J.-M. Mesnard, and B. Barbeau. HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated. *Retrovirology*, 3(1): 15, 2006.

L. B. Cook, M. Elemans, A. G. Rowan, and B. Asquith. HTLV-1: persistence and pathogenesis. *Virology*, 435(1):131–140, 2013.

S. Daenke, S. Nightingale, J. K. Cruickshank, and C. R. M. Bangham. Sequence variants of human T-cell lymphotropic virus type I from patients with tropical spastic paraparesis and adult T-cell leukemia do not distinguish neurological from leukemic isolates. *Journal of Virology*, 64(3):1278–1282, 1990.

S. Daenke, A. G. Kermode, S. E. Hall, G. Taylor, J. Weber, S. Nightingale, and C. R. M. Bangham. High activated and memory cytotoxic T-cell responses to HTLV-1 in healthy carriers and patients with tropical spastic paraparesis. *Virology*, 217(1):139–146, 1996.

O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

M. Elemans, N.-K. S. A. Basatena, and B. Asquith. The efficiency of the human CD8+ T cell response: how should we quantify it, what determines it, and does it matter? *PLoS Computational Biology*, 8(2), 2012.

S. A. Frank. *Immunology and Evolution of Infectious Disease*. Princeton University Press, 2002.

H. R. Fryer, J. Frater, A. Duda, M. G. Roberts, R. E. Phillips, A. R. McLean, S. T. Investigators, et al. Modelling the evolution and spread of HIV immune escape mutants. *PLoS Pathogens*, 6(11):e1001196–e1001196, 2010.

Y. Furukawa, M. Yamashita, K. Usuku, S. Izumo, M. Nakagawa, and M. Osame. Phylogenetic Subgroups of Human T Cell Lymphotropic Virus (HTLV) Type I in the tax Gene and Their Association with Different Risks for HTLV-IAssociated Myelopathy/Tropical Spastic Paraparesis. *Journal of Infectious Diseases*, 182(5):1343–1349, 2000.

Y. Furukawa, R. Kubota, M. Tara, S. Izumo, and M. Osame. Existence of escape mutant in HTLV-I *tax* during the development of adult T-cell leukemia. *Immunobiology*, 97(4): 987–993, 2001.

R. C. Gallo. The discovery of the first human retrovirus: HTLV-1 and HTLV-2. *Retrovirology*, 2(1):17–23, 2005.

G. Gaudray, F. Gachon, J. Basbous, M. Biard-Piechaczyk, C. Devaux, and J.-M. Mesnard. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. *Journal of Virology*, 76(24):12813–12822, 2002.

N. A. Gillet, N. Malani, A. Melamed, , N. Gormley, R. Carter, D. Bentley, C. Berry, F. D. Bushman, G. P. Taylor, and C. R. M. Bangham. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood*, 117(11): 3113–3122, 2011.

C. Gini. Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi*, 1, 1912.

H. Gómez-Acevedo and M. Y. Li. Backward bifurcation in a model for HTLV-I infection of $CD4^+$ T cells. *Bulletin of Mathematical Biology*, 67(1):101–114, 2005.

H. Gómez-Acevedo, M. Y. Li, and S. Jacobson. Multistability in a model for CTL response to HTLV-I infection and its implications to HAM/TSP development and prevention. *Bulletin of Mathematical Biology*, 72(3):681–696, 2010.

D. U. Gonçalves, F. A. Proietti, J. G. R. Ribas, M. G. Araújo, S. R. Pinheiro, A. C. Guedes, and A. B. F. Carneiro-Proietti. Epidemiology, treatment, and prevention of human T-cell leukemia virus type 1-associated diseases. *Clinical Microbiology Reviews*, 23(3):577–589, 2010.

M. Goodenow, T. Huet, W. Saurin, S. Kwok, J. Sninsky, and S. Wain-Hobson. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *Journal of Acquired Immune Deficiency Syndromes*, 2(4):344–352, 1989.

P. K. Goon, T. Igakura, E. Hanon, A. J. Mosley, B. Asquith, K. G. Gould, G. P. Taylor, J. N. Weber, and C. R. Bangham. High circulating frequencies of tumor necrosis factor alpha- and interleukin-2-secreting human T-lymphotropic virus type 1 (HTLV-1)-specific CD4+ T cells in patients with HTLV-1-associated neurological disease. *Journal of Virology*, 77 (17):9716–9722, 2003.

P. K. C. Goon, A. Biancardi, N. Fast, T. Igakura, E. Hanon, A. J. Mosley, B. Asquith, K. G. Gould, S. Marshall, G. P. Taylor, and C. R. M. Bangham. Human T cell lymphotropic virus (HTLV) type-1-specific $CD8^+$ T cells: frequency and immunodominance hierarchy. *Journal of Infectious Diseases*, 189(12):2294–2298, 2004a.

P. K. C. Goon, T. Igakura, E. Hanon, A. J. Mosley, A. Barfield, A. L. Barnard, L. Kaftantzi, Y. Tanaka, G. P. Taylor, J. N. Weber, and C. R. M. Bangham. Human T cell lymphotropic virus type 1 (HTLV-I)-specific $CD4^+$ T Cells: immunodominance hierarchy and preferential Infection with HTLV-I. *The Journal of Immunology*, 172(3):1735–1743, 2004b.

S. Gupta. Parasite immune escape: new views into host-parasite interactions. *Current Opinion in Microbiology*, 8(4):428–433, 2005.

S. Gupta and R. M. Anderson. Population structure of pathogens: the role of immune selection. *Parasitology Today*, 15(12):497–501, 1999.

S. Gupta and M. C. J. Maiden. Exploring the evolution of diversity in pathogen populations. *Trends in Microbiology*, 9(4):181–185, 2001.

S. Gupta, J. Swinton, and R. M. Anderson. Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 256(1347):231–238, 1994.

S. Gupta, M. C. J. Maiden, I. M. Feavers, S. Nee, R. M. May, and R. M. Anderson. The maintenance of strain structure in populations of recombining infectious agents. *Nature Medicine*, 2(4):437–442, 1996.

S. Gupta, N. Ferguson, and R. Anderson. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science*, 280(5365):912–915, 1998.

G. Hardin. The competitive exclusion principle. *Science*, 131(3409):1292–1297, 1960.

P. Höllsberg. Mechanisms of T-Cell activation by human T-cell lymphotropic virus type I. *Microbiology and Molecular Biology Reviews*, 63(2):308–333, 1999.

T. Igakura, J. C. Stinchcombe, P. K. C. Goon, G. P. Taylor, J. N. Weber, G. M. Griffiths, Y. Tanaka, M. Osame, and C. R. M. Bangham. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science*, 299(5613):1713–1716, 2003.

S. Jacobson. Immunopathogenesis of human T cell lymphotropic virus type I-associated neurologic disease. *Journal of Infectious Diseases*, 186(Supplement 2):S187–S192, 2002.

K. J. Jeffery, K. Usuku, S. E. Hall, W. Matsumoto, G. P. Taylor, J. Procter, M. Bunce, G. S. Ogg, K. I. Welsh, J. N. Weber, A. L. Lloyd, M. A. Nowak, M. Nagai, D. Kodama, S. Izumo, M. Osame, and C. R. M. Bangham. HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proceedings of the National Academy of Sciences*, 96(7):3848–3853, 1999.

K. J. Jeffery, A. A. Siddiqui, M. Bunce, A. L. Lloyd, A. M. Vine, A. D. Witkover, S. Izumo, K. Usuku, K. I. Welsh, M. Osame, and C. R. M. Bangham. The influence of HLA class I alleles and heterozygosity on the outcome of human T cell lymphotropic virus type I infection. *Journal of Immunology*, 165(12):7278–7284, 2000.

J. M. Johnson, R. Harrod, and G. Franchini. Molecular biology and pathogenesis of the human T-cell leukaemia/lymphotropic virus Type-1 (HTLV-1). *International Journal of Experimental Pathology*, 82(3):135–147, 2001.

D. Kalajdzievska and M. Y. Li. Modeling the effects of carriers on transmission dynamics of infectious diseases. *Mathematical Biosciences and Engineering*, 8(3):711–722, 2011.

M. Kannagi, S. Harada, I. Maruyama, H. Inoko, H. Igarashi, G. Kuwashima, S. Sato, M. Morita, M. Kidokoro, M. Sugimoto, et al. Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. *International Immunology*, 3(8):761–767, 1991.

T. Kattan, A. MacNamara, A. G. Rowan, H. Nose, A. J. Mosley, Y. Tanaka, G. P. Taylor, B. Asquith, and C. R. Bangham. The avidity and lytic efficiency of the CTL response to HTLV-1. *The Journal of Immunology*, 182(9):5723–5729, 2009.

T. Kinoshita, A. Tsujimoto, and K. Shimotohno. Sequence variations in ltr and env regions of HTLV-I do not discriminate between the virus from patients with HTLV-I-associated myelopathy and adult T-cell leukemia. *International Journal of Cancer*, 47(4):491–495, 1991.

D. Kirschner and G. F. Webb. A model for treatment strategy in the chemotherapy of AIDS. *Bulletin of Mathematical Biology*, 58(2):367–390, 1996.

A. Korobeinikov. Global properties of basic virus dynamics models. *Bulletin of Mathematical Biology*, 66(4):879–883, 2004.

A. Korobeinikov and G. C. Wake. Lyapunov functions and global stability for SIR, SIRS, and SIS epidemiological models. *Applied Mathematics Letters*, 15(8):955–960, 2002.

R. Kubota, T. Kawanishi, H. Matsubara, A. Manns, and S. Jacobson. HTLV-I specific IFN-γ+ CD8+ lymphocytes correlate with the proviral load in peripheral blood of infected individuals. *Journal of Neuroimmunology*, 102(2):208–215, 2000.

R. Kubota, Y. Furukawa, S. Izumo, K. Usuku, and M. Osame. Degenerate specificity of HTLV-1–specific CD8+ T cells during viral replication in patients with HTLV-1–associated myelopathy (HAM/TSP). *Blood*, 101(8):3074–3081, 2003.

R. Kubota, K. Hanada, Y. Furukawa, K. Arimura, M. Osame, T. Gojobori, and S. Izumo. Genetic stability of human T lymphotropic virus type I despite antiviral pressures by CTLs. *Journal of Immunology*, 178(9):5966–5972, 2007.

J. P. LaSalle. *The Stability of Dynamical Systems*. Number 25 in CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1976.

A. J. Leslie, K. J. Pfafferott, P. Chetty, R. Draenert, M. M. Addo, M. Feeney, Y. Tang, E. C. Holmes, T. Allen, J. G. Prado, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nature Medicine*, 10(3):282–289, 2004.

M. C. Levin and S. Jacobson. HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP): a chronic progressive neurologic disease associated with immunologically mediated damage to the central nervous system. *Journal of NeuroVirology*, 3(2):126–140, 1997.

M. Y. Li and A. G. Lim. Modelling the role of Tax expression in HTLV-I persistence in vivo. *Bulletin of Mathematical Biology*, 73(12):3008–3029, 2011.

A. G. Lim and P. K. Maini. HTLV-I infection: A dynamic struggle between viral persistence and host immunity. *Journal of Theoretical Biology*, 352:92–108, 2014.

C. A. Macken and A. S. Perelson. A multistage model for the action of cytotoxic T lymphocytes in multicellular conjugates. *The Journal of Immunology*, 132(4):1614–1624, 1984.

A. MacNamara, A. Rowan, S. Hilburn, U. Kadolsky, H. Fujiwara, K. Suemori, M. Yasukawa, G. Taylor, C. R. M. Bangham, and B. Asquith. HLA class I binding of HBZ determines outcome in HTLV-I infection. *PLoS Pathogens*, 6(9):e1001117, 2010.

M. Matsuoka. Human T-cell leukemia virus type I (HTLV-I) infection and the onset of adult T-cell leukemia (ATL). *Retrovirology*, 2(1):27, 2005.

M. Matsuoka and P. L. Green. The HBZ gene, a key player in HTLV-I pathogenesis. *Retrovirology*, 6(1):71, 2009.

K. N. Meekings, J. Leipzig, F. D. Bushman, G. P. Taylor, and C. R. M. Bangham. HTLV-1 integration into transcriptionally active genomic regions Is associated with proviral expression and with HAM/TSP. *PLoS Pathogens*, 4(3):e1000027, 2008.

A. Melamed, D. J. Laydon, N. A. Gillet, Y. Tanaka, G. P. Taylor, and C. R. M. Bangham. Genome-wide determinants of proviral targeting, clonal abundance and expression in natural HTLV-1 infection. *PLoS pathogens*, 9(3):e1003271, 2013.

F. Mortreux, A.-S. Gabet, and E. Wattel. Molecular and cellular aspects of HTLV-I associated leukemogenesis *in vivo*. *Leukemia*, 17(1):26–38, 2003.

A. J. Mosley and C. R. M. Bangham. A new hypothesis for the pathogenesis of human T-lymphotropic virus type 1 associated myelopathy/tropical spastic paraparesis. *Bioscience Hypotheses*, 2(3):118–124, 2009. ISSN 1756-2392.

A. J. Mosley, B. Asquith, and C. R. M. Bangham. Cell-mediated immune response to human T-lymphotropic virus type I. *Viral Immunology*, 18(2):293–305, 2005.

D. Moynet, J.-F. Pouliquen, D. Londos-Gagliardi, R. P. Buigues, J.-F. Moreau, I. Bedjabaga, M.-C. Georges, A. Talarmin, M. Joubert, H. Fleury, P. Vincendeau, and B. Guillemain. High variability of HTLV-I in a remote population of Gabon as compared to that of a similar population of French Guiana. *Virus Genes*, 23(3):257–261, 2001.

M. Nagai and M. Osame. Human T-cell lymphotropic virus type I and neurological diseases. *Journal of Neurovirology*, 9(2):228–235, 2003.

M. Nagai, K. Usuku, W. Matsumoto, D. Kodama, N. Takenouchi, T. Moritoyo, S. Hashiguchi, M. Ichinose, C. R. M. Bangham, S. Izumo, and M. Osame. Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. *Journal of Neurovirology*, 4 (6):586–593, 1998.

M. Nagai, R. Kubota, T. F. Greten, J. P. Schneck, T. P. Leist, and S. Jacobson. Increased activated human T cell lymphotropic virus type I (HTLV-I) Tax11-19-specific memory and effector CD8+ cells in patients with HTLV-I-associated myelopathy/tropical spastic paraparesis: correlation with HTLV-I provirus load. *Journal of Infectious Diseases*, 183 (2):197–205, 2001.

P. W. Nelson, J. D. Murray, and A. S. Perelson. A model of HIV-1 pathogenesis that includes an intracellular delay. *Mathematical Biosciences*, 163(2):201–215, 2000.

S. Niewiesk and C. R. M. Bangham. Evolution in a chronic RNA virus infection: selection on HTLV-I Tax protein differs between healthy carriers and patients with tropical spastic paraparesis. *Journal of Molecular Evolution*, 42:452–458, 1996.

S. Niewiesk, S. Daenke, C. E. Parker, G. Taylor, J. Weber, S. Nightingale, and C. R. M. Bangham. The transactivator gene of human T-cell leukemia virus type I is more variable within and between healthy carriers than patients with tropical spastic paraparesis. *Journal of Virology*, 68(10):6778–6781, 1994.

S. Niewiesk, S. Daenke, C. E. Parker, G. Taylor, J. Weber, S. Nightingale, and C. R. M. Bangham. Naturally occurring variants of human T-cell leukemia virus type I Tax protein impair its recognition by cytotoxic T lymphocytes and the transactivation function of Tax. *Journal of Virology*, 69(4):2649–2653, 1995.

M. A. Nowak. Immune responses against multiple epitopes: a theory for immunodominance and antigenic variation. volume 7, pages 83–92. Philadelphia, PA: Saunders Scientific Publications, 1996.

M. A. Nowak and C. R. M. Bangham. Population dynamics of immune responses to persistent viruses. *Science*, 272(5258):74–79, 1996.

M. A. Nowak, R. M. May, R. E. Phillips, S. Rowland-Jones, D. G. Lalloo, S. McAdam, P. Klenerman, B. Köppe, K. Sigmund, C. R. M. Bangham, and A. J. McMichael. Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature*, 375(6532): 606–611, 1995a.

M. A. Nowak, R. M. May, and K. Sigmund. Immune responses against multiple epitopes. *Journal of Theoretical Biology*, 175(3):325–353, 1995b.

P. Parham. *The Immune System*. Garland Science, 2005.

C. E. Parker, S. Daenke, S. Nightingale, and C. R. M. Bangham. Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. *Virology*, 188(2):628–636, 1992.

C. E. Parker, S. Nightingale, G. P. Taylor, J. Weber, and C. R. M. Bangham. Circulating anti-Tax cytotoxic T lymphocytes from human T-cell leukemia virus type I-infected people, with and without tropical spastic paraparesis, recognize multiple epitotes simultaneously. *Journal of Virology*, 68(5):2860–2868, 1994.

A. S. Perelson. Modeling the interaction of the immune system with HIV. *Lecture Notes in Biomathematics*, 83:350–370, 1989.

A. S. Perelson. Modelling viral and immune system dynamics. *Nature Reviews Immunology*, 2:28–36, 2002.

A. S. Perelson and P. W. Nelson. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM review*, 41(1):3–44, 1999.

A. S. Perelson, D. E. Kirschner, and R. D. Boer. Dynamics of HIV infection of CD4+ T cells. *Mathematical Biosciences*, 114(1):81–125, 1993.

C. Pique, A. Ureta-Vidal, A. Gessain, B. Chancerel, O. Gout, R. Tamouza, F. Agis, and M.-C. Dokhélar. Evidence for the chronic in vivo production of human T cell leukemia virus type I Rof and Tof proteins from cytotoxic T lymphocytes directed against viral peptides. *Journal of Experimental Medicine*, 191(3):567–572, 2000.

F. A. Proietti, A. B. F. Carneiro-Proietti, B. C. Catalan-Soares, and E. L. Murphy. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene*, 24:6058–6068, 2005.

A. Rambaut, D. Posada, K. A. Crandall, and E. C. Holmes. The causes and consequences of HIV evolution. *Nature Reviews Genetics*, 5(1):52–61, 2004.

M. Recker and S. Gupta. A model for pathogen population structure with cross-protection depending on the extent of overlap in antigenic variant repertoires. *Journal of Theoretical Biology*, 232(3):363–373, 2005.

M. Recker, S. Nee, P. C. Bull, S. Kinyanjui, K. Marsh, C. Newbold, and S. Gupta. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. *Nature*, 429(6991):555–558, 2004.

M. Recker, O. G. Pybus, S. Nee, and S. Gupta. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. *Proceedings of the National Academy of Sciences*, 104(18):7711–7716, 2007.

B. Renjifo, I. Borrero, and M. Essex. Tax mutation associated with tropical spastic paraparesis/human T-cell leukemia virus type I-associated myelopathy. *Journal of Virology*, 69(4):2611–2616, 1995.

R. M. Ribeiro, H. Mohri, D. D. Ho, and A. S. Perelson. *In vivo* dynamics of T cell activation, proliferation, and death in HIV-1 infection: Why are $CD4^+$ but not $CD8^+$ T cells depleted? *Proceedings of the National Academy of Sciences*, 99(24):15572–15577, 2002.

J. H. Richardson, P. Höllsberg, A. Windhagen, L. A. Child, and A. M. L. Lever. Variable immortalizing potential and frequent virus latency in blood-derived T-cell clones infected with human T-cell leukemia virus type I. *Blood*, 89(9):3303–3314, 1997.

A. G. Rowan and C. R. Bangham. Is there a role for HTLV-1-specific CTL in adult T-Cell leukemia/lymphoma? *Leukemia Research and Treatment*, 2012, 2011.

M. Saito, T. Matsuzaki, Y. Satou, J. ichirou Yasunaga, K. Saito, K. Arimura, M. Matsuoka, and Y. Ohara. *In vivo* expression of the HBZ gene of HTLV-I correlates with proviral load, inflammatory markers and disease severity in HTLV-I associated myelopathy/tropical spastic paraparesis (HAM/TSP). *Retrovirology*, 6:19, 2009.

Y. Satou, J. ichirou Yasunaga, M. Yoshida, and M. Matsuoka. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proceedings of the National Academy of Sciences*, 103(3):720–725, 2006.

N.-K. Seich al Basatena, A. MacNamara, A. M. Vine, C. L. Thio, J. Astemborski, K. Usuku, M. Osame, G. D. Kirk, S. M. Donfield, J. J. Goedert, C. R. M. Bangham, M. Carrington, S. I. Khakoo, and B. Asquith. KIR2DL2 enhances protective and detrimental HLA class I-mediated immunity in chronic viral infection. *PLoS Pathogens*, 7(10):e1002270, 2011.

O. J. Semmes and K.-T. Jeang. Mutational analysis of human T-cell leukemia virus type I Tax: regions necessary for function determined with 47 mutant proteins. *Journal of Virology*, 66(12):7183–7192, 1992.

H. Shiraki, Y. Sagara, and Y. Inoue. *Cell-to-cell transmission of HTLV-I*, pages 303–316. Japan Scientific Societies Press, 2003. Gann Monograph on Cancer Research, No. 50. Two Decades of Adult T-cell Leukemia and HTLV-I Research.

R. E. T. Smith, S. Niewiesk, S. Booth, C. R. M. Bangham, and S. Daenke. Functional conservation of HTLV-1 Rex balances the immune pressure for sequence variation in the *Rex* gene. *Virology*, 237(2):397–403, 1997.

R. P. Smyth, M. P. Davenport, and J. Mak. The origin of genetic diversity in HIV-1. *Virus Research*, 169(2):415–429, 2012.

K. Suemori, H. Fujiwara, T. Ochi, T. Ogawa, M. Matsuoka, T. Matsumoto, J.-M. Mesnard, and M. Yasukawa. HBZ is an immunogenic protein, but not a target antigen for human T-cell leukemia virus type 1-specific cytotoxic T lymphocytes. *Journal of General Virology*, 90:1806–1811, 2009.

U. Tomaru, Y. Yamano, and S. Jacobson. *HTLV-I infection and the nervous system*, chapter 23, pages 285–299. Oxford University Press, 2005. Clinical Neuroimmunology, 2nd Edition.

S. Ugolini, C. Arpin, N. Anfossi, T. Walzer, A. Cambiaggi, R. Förster, M. Lipp, R. E. M. Toes, C. J. Melief, J. Marvel, and E. Vivier. Involvement of inhibitory NKRs in the survival of a subset of memory-phenotype CD8+ T cells. *Nature immunology*, 2(5):430–435, 2001.

A. M. Vine, A. D. Witkover, A. L. Lloyd, K. J. Jeffery, A. Siddiqui, S. E. Marshall, M. Bunce, N. Eiraku, S. Izumo, K. Usuku, et al. Polygenic control of human T lymphotropic virus type I (HTLV-I) provirus load and the risk of HTLV-I-associated myelopathy/tropical spastic paraparesis. *Journal of Infectious Diseases*, 186(7):932–939, 2002.

E. Wattel, M. Cavrois, A. Gessain, and S. Wain-Hobson. Clonal expansion of infected cells: a way of life for HTLV-I. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 13:S92–S99, 1996.

P. S. Wikramaratna, M. Sandeman, M. Recker, and S. Gupta. The antigenic evolution of influenza: drift or thrift? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614), 2013.

D. Wodarz. Mathematical models of immune effector responses to viral infections: Virus control versus the development of pathology. *Journal of Computational and Applied Mathematics*, 184(1):301–319, 2005.

D. Wodarz, M. A. Nowak, and C. R. M. Bangham. The dynamics of HTLV-I and the CTL response. *Immunology Today*, 20(5):220–227, 1999.

D. Wodarz, S. E. Hall, K. Usuku, M. Osame, G. S. Ogg, A. J. McMichael, M. A. Nowak, and C. R. M. Bangham. Cytotoxic T-cell abundance and virus load in human immunodeficiency virus type 1 and human T-cell leukaemia virus type 1. *Proceedings of the Royal Society of London. Biological Sciences*, 268(1473):1215–1221, 2001.

D. Wodarz, J. P. Christensen, and A. R. Thomsen. The importance of lytic and nonlytic immune responses in viral infections. *Trends in Immunology*, 23(4):194–200, 2002.

S. Yashiki, T. Fujiyoshi, N. Arima, M. Osame, M. Yoshinaga, Y. Nagata, M. Tara, K. Nomura, A. Utsunomiya, S. Hanada, K. Tajima, and S. Sonoda. HLA-A* 26, HLA-B* 4002, HLA-B* 4006, and HLA-B* 4801 alleles predispose to adult T cell leukemia: the limited recognition of HTLV type 1 tax peptide anchor motifs and epitopes to generate anti-HTLV type 1 tax CD8+ cytotoxic T lymphocytes. *AIDS research and human retroviruses*, 17(11):1047–1061, 2001.

D. Zagury, J. Bernard, P. Jeannesson, N. Thiernesse, and J.-C. Cerottini. Studies on the Mechanism of T Cell-Mediated Lysis at the Single Effector Cell Level I. Kinetic Analysis of Lethal Hits and Target Cell Lysis in Multicellular Conjugates. *The Journal of Immunology*, 123(4):1604–1609, 1979.